PCT

ORLD INTELLECTUAL PROPERTY ORGANIZA' International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

(11) International Publication Number:

WO 99/32644

C12N 15/85, 9/10, C07K 16/30, 16/40, C12Q 1/68, A01K 67/027

(43) International Publication Date:

1 July 1999 (01.07.99)

(21) International Application Number:

PCT/IB98/02133

A2

(22) International Filing Date:

22 December 1998 (22.12.98)

(30) Priority Data:

08/996,306 60/099,658

US 22 December 1997 (22.12.97) 9 September 1998 (09.09.98) US

(71) Applicant (for all designated States except US): GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR).

(72) Inventors; and

(75) Inventors/Applicants (for US only): COHEN, Daniel [FR/FR]; 5, avenue Odette, F-94210 Fontenay-sous-Bois (FR). BLUMENFELD, Marta [FR/FR]; 5, rue Tagore, F-75013 Paris (FR). CHUMAKOV, Ilya [FR/FR]; 196, nue des-Chèvrefeuilles, F-77000-Vaux=te-Péñil (FR). BOUGUEL-EREF, Lydie [FR/FR]; 108, avenue Victor-Hugo, F-92170 Vanves (FR).

(74) Agents: MARTIN, Jean-Jacques et al.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: PROSTATE CANCER GENE

(57) Abstract

The present invention relates to PG1, a gene associated with prostate cancer. The invention provides polynucleotides including biallelic markers derived from PG1 and from flanking genomic regions. Primers hybridizing to these biallelic markers and regions flanking are also provided. This invention provides polynucleotides and methods suitable for genotyping a nucleic acid containing sample for one or more biallelic markers of the invention. Further, the invention provides methods to detect a statistical correlation between a biallelic marker allele and prostate cancer and between a haplotype and prostate cancer. The invention also relates to diagnostic methods of determining whether an individual is at risk for developing prostate cancer, and whether an individual suffers from prostate concer as a result of a mutation in the PG1 gene.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AM AT	Armenia	FR	France	LU	Luxembourg	SN	Senegal
	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AU	Australia Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
AZ	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BA		GH	Ghana	MG	Madagascar	ТJ	Tajikistan
BB	Barbados	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BE	Belgium	GR	Greece		Republic of Macedonia	TR	Turkey
BF	Burkina Faso	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BG	Bulgaria	IE	Ireland	MN	Mongolia	UA	Ukraine
BJ	Benin		Israel	MR	Mauritania	UG	Uganda
BR	Brazil	IL		MW	Malawi	US	United States of America
BY	Belarus	IS	Iceland	' MX	Mexico	UZ	Uzbekistan
CA	Canada	IT	Italy	NE NE	Niger	VN	Viet Nam
CF	Central African Republic	JP	Japan	NL NL	Netherlands	YU	Yugoslavia
CG	Congo	KE	Kenya			zw	Zimbabwe
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	Z	Zimbaowe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

PROSTATE CANCER GENE

Background of the Invention

5

10

25

A cancer is a clonal proliferation of cells produced as a consequence of cumulative genetic damage that finally results in unrestrained cell growth, tissue invasion and metastasis (cell transformation). Regardless of the type of cancer, transformed cells carry damaged DNA in many forms: as gross chromosomal translocations or, more subtly, as DNA amplification, rearrangement or even point mutations.

Some oncogenic mutations is inherited in the germline, thus predisposing the mutation carrier to an increased risk of cancer. However, in a majority of cases, cancer does not occur as a simple monogenic disease with clear Mendelian inheritance. There is only a two- or threefold increased risk of cancer among first-degree relatives for many cancers (Mulvihill JJ, Miller RW & Fraumeni JF, 1977, Genetics of human cancer Vol 3, New York Raven Press). Alternatively, DNA damage is 15 acquired somatically, probably induced by exposure to environmental carcinogens. mutations are generally responsible for the vast majority of cancer cases.

Studies of the age dependence of cancer have suggested that several successive mutations are needed to convert a normal cell into an invasive carcinoma. Since human mutation rates are typically 10⁻⁶/gene/cell, the chance of a single cell undergoing many independent mutations is very low (Loeb 20 LA, Cancer Res 1991, 51: 3075-3079). Cancer nevertheless happens because of a combination of two mechanisms. Some mutations enhance cell proliferation, increasing the target population of cells for the next mutation. Other mutations affect the stability of the entire genome, increasing the overall mutation rate, as in the case of mismatch repair proteins (reviewed in Arnheim N & Shibata D, Curr. Op. Genetics & Development, 1997, 7:364-370).

An intricate process known as the cell cycle drives normal proliferation of cells in an organism. Regulation of the extent of cell cycle activity and the orderly execution of sequential steps within the cycle ensure the normal development and homeostasis of the organism. Conversely, many of the properties of cancer cells - uncontrolled proliferation, increased mutation rate, abnormal translocations and gene amplifications - can be attributed directly to perturbations of the normal 30 regulation or progression of the cycle. In fact, many of the genes that have been identified over the past several decades as being involved in cancer, can now be appreciated in terms of their direct or indirect role in either regulating entry into the cell cycle or coordinating events within the cell cycle.

Recent studies have identified three groups of genes which are frequently mutated in cancer. The first group of genes, called oncogenes, are genes whose products activate cell proliferation. The 35 normal non-mutant versions are called protooncogenes. The mutated forms are excessively or inappropriately active in promoting cell proliferation, and act in the cell in a dominant way in that a

2

single mutant allele is enough to affect the cell phenotype. Activated oncogenes are rarely transmitted as germline mutations since they may probably be lethal when expressed in all the cells. Therefore oncogenes can only be investigated in tumor tissues.

Oncogenes and protooncogenes can be classified into several different categories according to their function. This classification includes genes that code for proteins involved in signal transduction such as: growth factors (i.e., sis, int-2); receptor and non-receptor protein-tyrosine kinases (i.e., erbB, src, bcr-abl, met, trk); membrane-associated G proteins (i.e., ras); cytoplasmic protein kinases (i.e., mitogen-activated protein kinase –MAPK- family, raf, mos, pak), or nuclear transcription factors (i.e., myc, myb, fos, jun, rel) (for review see Hunter T, 1991 Cell 64:249; Fanger GR et al., 1997 Curr.Op.Genet.Dev.7:67-74; Weiss FU et al., ibid. 80-86).

The second group of genes which are frequently mutated in cancer, called tumor suppressor genes, are genes whose products inhibit cell growth. Mutant versions in cancer cells have lost their normal function, and act in the cell in a recessive way in that both copies of the gene must be inactivated in order to change the cell phenotype. Most importantly, the tumor phenotype can be rescued by the wild type allele, as shown by cell fusion experiments first described by Harris and colleagues (Harris H et al.,1969,Nature 223:363-368). Germline mutations of tumor suppressor genes is transmitted and thus studied in both constitutional and tumor DNA from familial or sporadic cases. The current family of tumor suppressors includes DNA-binding transcription factors (i.e., p53, WT1), transcription regulators (i.e., RB, APC, probably BRCA1), protein kinase inhibitors (i.e., p16), among others (for review, see Haber D & Harlow E, 1997, Nature Genet. 16:320-322).

The third group of genes which are frequently mutated in cancer, called mutator genes, are responsible for maintaining genome integrity and/or low mutation rates. Loss of function of both alleles increase cell mutation rates, and as consequence, proto-oncogenes and tumor suppressor genes is mutated. Mutator genes can also be classified as tumor suppressor genes, except for the fact that tumorigenesis caused by this class of genes cannot be suppressed simply by restoration of a wild-type allele, as described above. Genes whose inactivation may lead to a mutator phenotype include mismatch repair genes (i.e., MLH1, MSH2), DNA helicases (i.e., BLM, WRN) or other genes involved in DNA repair and genomic stability (i.e., p53, possibly BRCA1 and BRCA2) (For review see Haber D & Harlow E, 1997, Nature Genet. 16:320-322; Fishel R & Wilson T. 1997, Curr.Op.Genet.Dev.7: 105-113; Ellis NA,1997 ibid.354-363).

The recent development of sophisticated techniques for genetic mapping has resulted in an ever expanding list of genes associated with particular types of human cancers. The human haploid genome contains an estimated 80,000 to 100,000 genes scattered on a 3 x 10⁹ base-long double-stranded DNA. Each human being is diploid i.e., possesses two haploid genomes, one from paternal origin, the other from maternal origin. The sequence of a given genetic locus may vary between individuals in a population or between the two copies of the locus on the chromosomes of a single

individual. Genetic mapping techniques often exploit these differences, which are called polymorphisms, to map the location of genes associated with human phenotypes.

3

One mapping technique, called the loss of heterozygosity (LOH) technique, is often employed to detect genes in which a loss of function results in a cancer, such as the tumor suppressor genes described above. Tumor suppressor genes often produce cancer via a two hit mechanism in which a first mutation, such as a point mutation (or a small deletion or insertion) inactivates one allele of the tumor suppressor gene. Often, this first mutation is inherited from generation to generation.

A second mutation, often a spontaneous somatic mutation such as a deletion which deletes all or part of the chromosome carrying the other copy of the tumor suppressor gene, results in a cell in which both copies of the tumor suppressor gene are inactive.

As a consequence of the deletion in the tumor suppressor gene, one allele is lost for any genetic marker located close to the tumor suppressor gene. Thus, if the patient is heterozygous for a marker, the tumor tissue loses heterozygosity, becoming homozygous or hemizygous. This loss of heterozygosity generally provides strong evidence for the existence of a tumor suppressor gene in the lost region.

By genotyping pairs of blood and tumor samples from affected individuals with a set of highly polymorphic genetic markers, such as microsatellites, covering the whole genome, one can discover candidate locations for tumor suppressor genes. Due to the presence of contaminant non-tumor tissue in most pathological tumor samples, a decreased relative intensity rather than total loss of heterozygosity of informative microsatellites is observed in the tumor samples. Therefore, classic LOH analysis generally requires quantitative PCR analysis, often limiting the power of detection of this technique. Another limitation of LOH studies resides on the fact that they only allow the definition of rather large candidate regions, typically spanning over several megabases. Refinement of such candidate regions requires the definition of the minimally overlapping portion of LOH regions identified in tumor tissues from several hundreds of affected patients.

Another approach to genetic mapping, called linkage analysis, is based upon establishing a correlation between the transmission of genetic markers and that of a specific trait throughout generations within a family. In this approach, all members of a series of affected families are genotyped with a few hundred markers, typically microsatellite markers, which are distributed at an average density of one every 10 Mb. By comparing genotypes in all family members, one can attribute sets of alleles to parental haploid genomes (haplotyping or phase determination). The origin of recombined fragments is then determined in the offspring of all families. Those that co-segregate with the trait are tracked. After pooling data from all families, statistical methods are used to determine the likelihood that the marker and the trait are segregating independently in all families. As a result of the statistical analysis, one or several regions are selected as candidates, based on their high probability to carry a trait causing allele. The result of linkage analysis is considered as significant

PCT/IB98/02133 WO 99/32644

when the chance of independent segregation is lower than 1 in 1000 (expressed as a LOD score > 3). Identification of recombinant individuals using additional markers allows further delineation of the candidate linked region, which most usually ranges from 2 to 20 Mb.

Linkage analysis studies have generally relied on the use of microsatellite markers (also 5 called simple tandem repeat polymorphisms, or simple sequence length polymorphisms). These include small arrays of tandem repeats of simple sequences (di- tri- tetra- nucleotide repeats), which exhibit a high degree of length polymorphism, and thus a high level of informativeness. To date, only just more than 5,000 microsatellites have been ordered along the human genome (Dib et al., Nature 1996, 380: 152), thus limiting the maximum attainable resolution of linkage analysis to ca. 600 kb on average.

Linkage analysis has been successfully applied to map simple genetic traits that show clear Mendelian inheritance patterns. About 100 pathological trait-causing genes were discovered by linkage analysis over the last 10 years.

However, linkage analysis approaches have proven difficult for complex genetic traits, those 15 probably due to the combined action of multiple genes and/or environmental factors. In such cases, too large an effort and cost are needed to recruit the adequate number of affected families required for applying linkage analysis to these situations, as recently discussed by Risch, N. and Merikangas, K. (Science 1996, 273: 1516-1517). Finally, linkage analysis cannot be applied to the study of traits for which no available large informative families are available. Typically, this will be the case in any 20 attempt to identify trait-causing alleles involved in sporadic cases.

The incidence of prostate cancer has dramatically increased over the last decades. It averages 30-50/100,000 males both in Western European countries as well as within the US White male population. In these countries, it has recently become the most commonly diagnosed malignancy, being one of every four cancers diagnosed in American males. Prostate cancer's incidence is very much population specific, since it varies from 2/100,000 in China, to over 80/100,000 among African-American males.

In France, the incidence of prostate cancer is 35/100,000 males and it is increasing by 10/100,000 per decade. Mortality due to prostate cancer is also growing accordingly. It is the second cause of cancer death among French males, and the first one among French males aged over 70. This makes prostate cancer a serious burden in terms of public health, especially in view of the aging of populations.

An average 40% reduction in life expectancy affects males with prostate cancer. completely localized, prostate cancer can be cured by surgery, with however an average success rate of only ca. 50%. If diagnosed after metastasis from the prostate, prostate cancer is a fatal disease for which there is no curative treatment.

Early-stage diagnosis relies on Prostate Specific Antigen (PSA) dosage, and would allow the

WO 99/32644 PCT/IB98/02133 5

detection of prostate cancer seven years before clinical symptoms become apparent. The effectiveness of PSA dosage diagnosis is however limited, due to its inability to discriminate between malignant and non-malignant affections of the organ.

Therefore, there is a strong need for both a reliable diagnostic procedure which would enable early-stage prostate cancer prognosis, and for preventive and curative treatments of the disease. The present invention relates to the PG1 gene, a gene associated with prostate cancer, as well as diagnostic methods and reagents for detecting alleles of the gene which may cause prostate cancer, and therapies for treating prostate cancer.

10 Summary of the Invention

The present invention relates to the identification of a gene associated with prostate cancer, identified as the PG1 gene, and reagents, diagnostics, and therapies related thereto. The present invention is also based on the discovery of a novel set of PG1-related biallelic markers. See the definition of PG1-related biallelic markers in the Detailed Description Section. These markers are located in the coding regions as well as non-coding regions adjacent to the PG1 gene. The position of these markers and knowledge of the surrounding sequence has been used to design polynucleotide compositions which are useful in determining the identity of nucleotides at the marker position, as well as more complex association and haplotyping studies which are useful in determining the genetic basis for diseases including cancer and prostate cancer. In addition, the compositions and methods of the invention find use in the identification of the targets for the development of pharmaceutical agents and diagnostic methods, as well as the characterization of the differential efficacious responses to and side effects from pharmaceutical agents acting on diseases including cancer and prostate cancer.

A first embodiment of the invention is a recombinant, purified or isolated polynucleotide comprising, or consisting of a mammalian genomic sequence, gene, or fragments thereof. In one aspect the sequence is derived from a human, mouse or other mammal. In a preferred aspect, the genomic sequence of SEQ ID NO: 179 or the complement thereto. In a second preferred aspect, the genomic sequence is selected from one of the two mouse genomic fragments of SEQ ID NO: 182 and 183. In yet another aspect of this embodiment, the nucleic acid comprises nucleotides 1629 through 1870 of the sequence of SEQ ID NO: 179. Optionally, said polynucleotide consists of, consists essentially of, or comprises a contiguous span of nucleotides of a mammalian genomic sequence, preferably a sequence selected the following SEQ ID NOs: 179, 182, and 183, wherein said contiguous span is at least 6, 8, 10, 12, 15, 20, 25, 30, 50, 100, 200, or 500 nucleotides in length.

A second embodiment of the present invention is a recombinant, purified or isolated polynucleotide comprising, or consisting of a mammalian cDNA sequence, or fragments thereof. In one aspect the sequence is derived from a human, mouse or other mammal. In a preferred aspect, the

cDNA sequence is selected from the human cDNA sequences of SEQ ID NO: 3, 69, 112-124 or the complement thereto. In a second preferred aspect, the cDNA sequence is the mouse cDNA sequence of SEQ ID NO: 184. Optionally, said polynucleotide consists of, consists essentially of, or comprises a contiguous span of nucleotides of a mammalian genomic sequence, preferably a sequence selected the following SEQ ID NOs: 3, 69, 112-124 and 184, wherein said contiguous span is at least 6, 8, 10, 12, 15, 20, 25, 30, 50, 100, 200, or 500 nucleotides in length.

A third embodiment of the present invention is a recombinant, purified or isolated polynucleotide, or the complement thereof, encoding a mammalian PG1 protein, or a fragment thereof. In one aspect the PG1 protein sequence is from a human, mouse or other mammal. In a preferred aspect, the PG1 protein sequence is selected from the human PG1 protein sequences of SEQ ID NO: 4, 5, 70, and 125-136. In a second preferred aspect, the PG1 protein sequence is the mouse PG1 protein sequences of SEQ ID NO: 74. Optionally, said fragment of PG1 polypeptide consists of, consists essentially of, or comprises a contiguous stretch of at least 8, 10, 12, 15, 20, 25, 30, 50, 100 or 200 amino acids from SEQ ID NOs: 4, 5, 70, 74, and 125-136, as well as any other human, mouse or mammalian PG1 polypeptide.

A fourth embodiment of the invention are the polynucleotide primers and probes disclosed herein

A fifth embodiment of the present invention is a recombinant, purified or isolated polypeptide comprising or consisting of a mammalian PG1 protein, or a fragment thereof. In one aspect the PG1 protein sequence is from a human, mouse or other mammal. In a preferred aspect, the PG1 protein sequence is selected from the human PG1 protein sequences of SEQ ID NO: 4, 5, 70, and 125-136. In a second preferred aspect, the PG1 protein sequence is the mouse PG1 protein sequences of SEQ ID NO: 74. Optionally, said fragment of PG1 polypeptide consists of, consists essentially of, or comprises a contiguous stretch of at least 8, 10, 12, 15, 20, 25, 30, 50, 100 or 200 amino acids from SEQ ID NOs: 4, 5, 70, 74, and 125-136, as well as any other human, mouse or mammalian PG1 polypeptide.

A sixth embodiment of the present invention is an antibody composition capable of specifically binding to a polypeptide of the invention. Optionally, said antibody is polyclonal or monoclonal. Optionally, said polypeptide is an epitope-containing fragment of at least 8, 10, 12, 15, 20, 25, or 30 amino acids of a human, mouse, or mammalian PG1 protein, preferably a sequence selected from SEQ ID NOs: 4, 5, 70, 74, or 125-136.

A seventh embodiment of the present invention is a vector comprising any polynucleotide of the invention. Optionally, said vector is an expression vector, gene therapy vector, amplification vector, gene targeting vector, or knock-out vector.

35 An eighth embodiment of the present invention is a host cell comprising any vector of the invention.

A ninth embodiment of the present invention is a mammalian host cell comprising a PG1 gene disrupted by homologous recombination with a knock out vector.

A tenth embodiment of the present invention is a nonhuman host mammal or animal comprising a vector of the invention.

A further embodiment of the present invention is a nonhuman host mammal comprising a PG1 gene disrupted by homologous recombination with a knock out vector.

Another embodiment of the present invention is a method of determining whether an individual is at risk of developing cancer or prostate cancer at a later date or whether the individual suffers from cancer or prostate cancer as a result of a mutation in the PG1 gene comprising obtaining a nucleic acid sample from the individual; and determining whether the nucleotides present at one or more of the PG1-related biallelic markers of the invention are indicative of a risk of developing prostate cancer at a later date or indicative of prostate cancer resulting from a mutation in the PG1 gene. Optionally, said PG1-related biallelic is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66.

Another embodiment of the present invention is a method of determining whether an individual is at risk of developing prostate cancer at a later date or whether the individual suffers from prostate cancer as a result of a mutation in the PG1 gene comprising obtaining a nucleic acid sample from the individual and determining whether the nucleotides present at one or more of the polymorphic bases in a PG1-related biallelic marker. Optionally, said PG1-related biallelic is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66.

Another embodiment of the present invention is a method of obtaining an allele of the PG1 gene which is associated with a detectable phenotype comprising obtaining a nucleic acid sample from an individual expressing the detectable phenotype, contacting the nucleic acid sample with an agent capable of specifically detecting a nucleic acid encoding the PG1 protein, and isolating the nucleic acid encoding the PG1 protein. In one aspect of this method, the contacting step comprises contacting the nucleic acid sample with at least one nucleic acid probe capable of specifically hybridizing to said nucleic acid encoding the PG1 protein. In another aspect of this embodiment, the contacting step comprises contacting the nucleic acid sample with an antibody capable of specifically

30

binding to the PG1 protein. In another aspect of this embodiment, the step of obtaining a nucleic acid sample from an individual expressing a detectable phenotype comprises obtaining a nucleic acid sample from an individual suffering from prostate cancer.

Another embodiment of the present invention is a method of obtaining an allele of the PG1 gene which is associated with a detectable phenotype comprising obtaining a nucleic acid sample from an individual expressing the detectable phenotype, contacting the nucleic acid sample with an agent capable of specifically detecting a sequence within the 8p23 region of the human genome, identifying a nucleic acid encoding the PG1 protein in the nucleic acid sample, and isolating the nucleic acid encoding the PG1 protein. In one aspect of this embodiment, the nucleic acid sample is obtained from an individual suffering from cancer or prostate cancer.

Another embodiment of the present invention is a method of categorizing the risk of prostate cancer in an individual comprising the step of assaying a sample taken from the individual to determine whether the individual carries an allelic variant of PG1 associated with an increased risk of prostate cancer. In one aspect of this embodiment, the sample is a nucleic acid sample. In another 15 aspect a nucleic acid sample is assayed by determining the frequency of the PG1 transcripts present. In another aspect of this embodiment, the sample is a protein sample. In another aspect of this embodiment, the method further comprises determining whether the PG1 protein in the sample binds an antibody specific for a PG1 isoform associated with prostate cancer.

Another embodiment of the present invention is a method of categorizing the risk of prostate 20 cancer in an individual comprising the step of determining whether the identities of the polymorphic bases of one or more biallelic markers which are in linkage disequilibrium with the PG1 gene are indicative of an increased risk of prostate cancer.

Another embodiment of the present invention comprises a method of identifying molecules which specifically bind to a PG1 protein, preferably the protein of SEQ ID NO:4 or a portion thereof: comprising the steps of introducing a nucleic a nucleic acid encoding the protein of SEQ ID NO:4 or a portion thereof into a cell such that the protein of SEQ ID NO:4 or a portion thereof contacts proteins expressed in the cell and identifying those proteins expressed in the cell which specifically interact with the protein of SEQ ID NO:4 or a portion thereof.

Another embodiment of the present invention is a method of identifying molecules which specifically bind to the protein of SEQ ID NO: 4 or a portion thereof. One step of the method comprises linking a first nucleic acid encoding the protein of SEQ ID NO:4 or a portion thereof to a first indicator nucleic acid encoding a first indicator polypeptide to generate a first chimeric nucleic acid encoding a first fusion protein. The first fusion protein comprises the protein of SEQ ID NO:4 or a portion thereof and the first indicator polypeptide. Another step of the method comprises linking a 35 second nucleic acid nucleic acid encoding a test polypeptide to a second indicator nucleic acid encoding a second indicator polypeptide to generate a second chimeric nucleic acid encoding a second

fusion protein. The second fusion protein comprises the test polypeptide and the second indicator polypeptide. Association between the first indicator protein and the second indicator protein produces a detectable result. Another step of the method comprises introducing the first chimeric nucleic acid and the second chimeric nucleic acid into a cell. Another step comprises detecting the detectable result.

A further embodiment of the invention is a purified or isolated mammalian PG1 gene or cDNA sequence.

Further embodiments of the present invention include the nucleic acid and amino acid sequences of mutant or low frequency PG1 alleles derived from prostate cancer patients, tissues or 10 cell lines. The present invention also encompasses methods which utilize detection of these mutant PG1 sequences in an individual or tissue sample to diagnosis prostate cancer, assess the risk of developing prostate cancer or assess the likely severity of a particular prostate tumor.

Another embodiment of the invention encompasses any polynucleotide of the invention attached to a solid support. In addition, the polynucleotides of the invention which are attached to a 15 solid support encompass polynucleotides with any further limitation described in this disclosure, or those following: Optionally, said polynucleotides is specified as attached individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the inventions to a single solid support. Optionally, polynucleotides other than those of the invention may attached to the same solid support as polynucleotides of the invention. Optionally, when multiple polynucleotides are attached to a solid support they are attached at random locations, or in an ordered array. Optionally, said ordered array is addressable.

An additional embodiment of the invention encompasses the use of any polynucleotide for, or any polynucleotide for use in, determining the identity of an allele at a PG1-related biallelic marker. In addition, the polynucleotides of the invention for use in determining the identity of an allele at a 25 PG1-related biallelic marker encompass polynucleotides with any further limitation described in this disclosure, or those following: Optionally, said PG1-related biallelic marker is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-30 61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, said polynucleotide may comprise a sequence disclosed in the present specification. Optionally, said polynucleotide may consist of, or consist essentially of any polynucleotide described in the present specification. Optionally, said determining is performed in a hybridization assay, sequencing assay, microsequencing assay, or allele-specific amplification assay. Optionally, said polynucleotide is attached to a solid support, array, or addressable array. Optionally, said polynucleotide is labeled.

5

Another embodiment of the invention encompasses the use of any polynucleotide for, or any polynucleotide for use in, amplifying a segment of nucleotides comprising an PG1-related biallelic marker. In addition, the polynucleotides of the invention for use in amplifying a segment of nucleotides comprising a PG1-related biallelic marker encompass polynucleotides with any further limitation described in this disclosure, or those following: Optionally, said PG1-related biallelic marker is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, said polynucleotide may comprise a sequence disclosed in the present specification. Optionally, said polynucleotide may consist of, or consist essentially of any polynucleotide described in the present specification. Optionally, said amplifying is performed by a PCR or LCR. Optionally, said polynucleotide is attached to a solid support, array, or addressable array. Optionally, said polynucleotide is labeled.

A further embodiment of the invention encompasses methods of genotyping a biological sample comprising determining the identity of an allele at an PG1-related biallelic marker. In addition, the genotyping methods of the invention encompass methods with any further limitation described in this disclosure, or those following: Optionally, said PG1-related biallelic marker is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, said method further 25 comprises determining the identity of a second allele at said biallelic marker, wherein said first allele and second allele are not base paired (by Watson & Crick base pairing) to one another. Optionally, said biological sample is derived from a single individual or subject. Optionally, said method is performed in vitro. Optionally, said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome. Optionally, said biological sample is derived from multiple subjects or individuals. Optionally, said method further comprises amplifying a portion of said sequence comprising the biallelic marker prior to said determining step. Optionally, wherein said amplifying is performed by PCR, LCR, or replication of a recombinant vector comprising an origin of replication and said portion in a host cell. Optionally, wherein said determining is performed by a hybridization assay, sequencing assay, microsequencing assay, or allele-specific amplification assay.

An additional embodiment of the invention comprises methods of estimating the frequency of an allele in a population comprising determining the proportional representation of an allele at a PG1-

35

related biallelic marker in said population. In addition, the methods of estimating the frequency of an allele in a population of the invention encompass methods with any further limitation described in this disclosure, or those following: Optionally, said PG1-related biallelic marker is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, determining the proportional representation of an allele at a PG1-related biallelic marker is accomplished by determining the identity of the alleles for both copies of said biallelic marker present in the genome of each individual in said population and calculating the proportional representation of said allele at said PG1-related biallelic marker for the population. Optionally, determining the proportional representation is accomplished by performing a genotyping method of the invention on a pooled biological sample derived from a representative number of individuals, or each individual, in said population, and

A further embodiment of the invention comprises methods of detecting an association between a genotype and a phenotype, comprising the steps of a) genotyping at least one PG1-related biallelic marker in a trait positive population according to a genotyping method of the invention; b) genotyping said PG1-related biallelic marker in a control population according to a genotyping 20 method of the invention; and c) determining whether a statistically significant association exists between said genotype and said phenotype. In addition, the methods of detecting an association between a genotype and a phenotype of the invention encompass methods with any further limitation described in this disclosure, or those following: Optionally, said PG1-related biallelic marker is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected 25 from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, said control population is a trait negative population, or a random population. Optionally, each of said genotyping steps a) and b) is performed on a single pooled biological sample derived from each of said populations. Optionally, each of said genotyping of steps a) and b) is performed separately on biological samples derived from each individual in said population or a subsample thereof. Optionally, said phenotype is a disease, cancer or prostate cancer; a response to an anti-cancer agent or an anti-prostate cancer agent; or a side effect to an anti-cancer or anti-prostate cancer agent. 35 Optionally, said method comprises the additional steps of determining the phenotype in said trait positive and said control populations prior to step c).

12

An additional embodiment of the present invention encompasses methods of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising the steps of: a) genotyping at least one PG1-related biallelic marker for both copies of said set of biallelic marker present in the genome of each individual in said population or a subsample thereof, according to a genotyping method of the invention; b) genotyping a second biallelic marker by determining the identity of the allele at said second biallelic marker for both copies of said second biallelic marker present in the genome of each individual in said population or said subsample, according to a genotyping method of the invention; and c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency. In 10 addition, the methods of estimating the frequency of a haplotype of the invention encompass methods with any further limitation described in this disclosure, or those following: Optionally, said PG1related biallelic marker is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-15 600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, said second biallelic marker is a PG1-related biallelic marker; a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-20 73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, said PG1-related biallelic marker and said second biallelic marker are 4-77/151 and 4-66/145. Optionally, said haplotype determination method is an expectation-maximization algorithm.

An additional embodiment of the present invention encompasses methods of detecting an association between a haplotype and a phenotype, comprising the steps of: a) estimating the frequency of at least one haplotype in a trait positive population, according to a method of the invention for estimating the frequency of a haplotype; b) estimating the frequency of said haplotype in a control population, according to a method of the invention for estimating the frequency of a 30 haplotype; and c) determining whether a statistically significant association exists between said haplotype and said phenotype. In addition, the methods of detecting an association between a haplotype and a phenotype of the invention encompass methods with any further limitation described in this disclosure, or those following: Optionally, said PG1-related biallelic is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-

10

61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66. Optionally, said PG1-related biallelic marker and said second biallelic marker are 4-77/151 and 4-66/145. Optionally, said haplotype exhibits a p-value of < 1x 10⁻³ in an association with a trait positive population with cancer, preferably prostate cancer. Optionally, said control population is a trait negative population, or a random population. Optionally, said phenotype is a disease, cancer or prostate cancer; a response to an anti-cancer agent or an anti-prostate cancer agent, or a side effects to an anti-cancer or anti-prostate cancer agent. Optionally, said method comprises the additional steps of determining the phenotype in said trait positive and said control populations prior to step c).

Additional embodiments and aspects of the present invention are set forth in the Detailed Description of the Invention and the Examples.

Brief Description of the Drawings

Figure 1 is a diagram showing the BAC contig containing the PG1 gene and the positions of biallelic markers along the contig.

Figure 2 is a graph showing the results of the first screening of a prostate cancer association study and the significance of various biallelic markers as measured by their chi squared and p-values for a low density set of markers.

Figure 3 is a graph showing the results of the first screening of a prostate cancer association study and the significance of various biallelic markers as measured by their chi squared and p-values for a higher density set of markers.

Figure 4 is a table demonstrating the results of an haplotype analysis. Among all the theoretical potential different haplotypes based on 2 to 9 markers, 11 haplotypes showing a strong association with prostate cancer were selected, and their haplotype analysis results are shown here.

Figure 5 is a bar graph demonstrating the results of an experiment evaluating the significance (p-values) of the haplotype analysis shown in Figure 4.

Figure 6A is a table listing the biallelic markers used in the haplotype analysis of Figure 4. Figure 6B is a table listing additional biallelic markers in linkage disequilibrium with the PG1 gene.

Figure 7 is a table listing the positions of exons, splice sites, a stop codon, and a poly A site in 30 the PG1 gene.

Figure 8A is a diagram showing the genomic structure of PG1 in comparison with its most abundant mRNA transcript. Figure 8B is a more detailed diagram showing the genomic structure of PG1, including exons and introns.

Figure 9 is a table listing some of the homologies between the PG1 protein and known 35 \tilde{p}_1 oteins.

Figure 10 is a half-tome reproduction of a fluorescence micrograph of the perinuclear/nuclear

expression of PG1 in tumoral (PC3) and normal prostatic cell lines (PNT2). Vector "PG1": includes all the coding exons from exon 1 to 8. For PC3 (upper panel) and PNT2 (lower panel), the nucleus was labelled with Propidium iodide (IP, left panel). Note that EGFP fluorescence was detected in and around the nucleus (GFP, middle panel), as shown when the two pictures were overlapped (right panel).

Figure 11 is a half-tome reproduction of a fluorescence micrograph of the perinuclear/nuclear expression of PG1/1-4 in tumoral (PC3) and normal prostatic cell lines (PNT2). Vector "PG1/1-4" corresponds to an alternative messenger which is due to an alternative splicing, joining exon 1 to exon 4, and resulting in the absence of exons 2 and 3. For PC3 (upper panel) and PNT2 (lower panel), the nucleus was labelled with Propidium iodide (IP, left panel). Note that EGFP fluorescence was detected in and around the nucleus (GFP, middle panel), as shown when the two pictures were overlapped (right panel).

Figure 12 is a half-tome reproduction of a fluorescence micrograph of the perinuclear/nuclear expression of PG1/1-5 in tumoral prostatic cell line (PC3) and cytoplasmic expression of PG1/1-5 in normal prostatic cell line (PNT2). Vector "PG1/1-5" corresponds to an alternative messenger which is due to an alternative splicing, joining exon 1 to exon 5, and resulting in the absence of exons 2, 3 and 4. For PC3 (upper panel) and PNT2 (lower panels), the nucleus was labelled with Propidium iodide (IP). Note that in PC3 cells, EGFP fluorescence was detected in and around the nucleus (GFP, upper middle panel), as shown when the two picture were overlapped (upper right panel). In PNT2A cells, EGFP fluorescence was detected in the cytoplasm (GFP, lower left panel), as shown when the two pictures were overlapped (lower right panel).

Figure 13 is a half-tome reproduction of a fluorescence micrograph of the perinuclear/nuclear expression of a mutated form PG1 (PG1mut229) in normal prostatic cell line (PNT2). Vector "PG1/1-7" includes exons 1 to 6, and corresponds to the mutated form identified in genomic DNA of the prostatic tumoural cell line LNCaP. The nucleus was labelled with Propidium iodide (IP, left panel). EGFP fluorescence was detected in the cytoplasm (GFP, middle panel), as shown when the two pictures were overlapped (lower right panel).

Figure 14 is a diagram of the structure of the 14 alternative splice species found for human PG1 by the exons present. An * indicates that there is a stop codon in frame at that location. An arrow to the right at the right-hand side of a splice species indicates that the open-reading frame continues off of the chart. a space between exons indicates that the exon(s) is missing from that particular alternative splice species. An up arrow indicates that either exon 1 bis, 3 bis, or 5 bis has been inserted depending upon which is indicated. A bracket notation in exon 6, over an exon 6 bis notation indicates that the first 60 bases is missing from exon 6, and exon 6 bis is therefor present as a truncated form of exon 6.

Figure 15 is a table listing the results of a series of RT-PCR experiments that were performed

on RNA of normal prostate, normal prostatic cell lines (PNT1A, PNT1B and PNT2), and tumoral prostatic cell lines (LnCaPFCG, LnCaPJMB, CaHPV, Du145, PC3, and prostate tumors (ECP5 to ECP24) using all the possible combinations of primers (SEQ ID NOs: 137-178) specific to all of the possible splice junctions or exon borders in human PG1. An NT indicates that the experiment was not performed. An [+] indicates the use of an alternative splice species with exons 1, 3, 4, 7, and 8.

Figure 16 is a graph showing the results of association studies using markers spanning the 650 kb region of the 8p23 locus around PG1, using both single point analysis and haplotyping studies.

Figure 17 is a graph showing an enlarged view of the single point association results within a 160 kb region comprising the PG1 gene.

Figure 18A is a graph showing an enlarged view of the single point association results of 40 kb within the PG1 gene. Figure 18B is a table listing the location of markers within PG1 gene, the two possible alleles at each site. For each marker, the disease-associated allele is indicated first; its frequencies in cases and controls as well as the difference between both are shown; the odd-ratio and the p-value of each individual marker association are also shown.

Figure 19A is a table showing the results of a haplotype analysis study using 4 markers (marker Nos. 4-14, 99-217, 4-66 and 99-221)) within the 160 kb region shown in Figure 17. Figure 19B is a table showing the segmented haplotyping results according to the subject's age, and whether the prostate cancer cases were sporadic or familial, using the same markers 4 markers and the same individuals as were used to generate the results in Figure 19A.

Figure 20 is a table listing the haplotyping results and odd ratios for combinations of the 7 markers (99-622; 4-77; 4-71; 4-73; 99-598; 99-576; 4-66) within PG1 gene that were shown in Figure 18 to have p-values more significant than 1.10⁻². All of the 2-, 3-, 4-, 5-, 6- and 7-marker haplotypes were tested.

Figure 21 is a graph showing the distribution of statistical significance, as measured by Chisquare values, for each series of possible x-marker haplotypes, (x = 2, 3 or 4) using all of the 19 markers listed in Figure 18B.

Detailed Description of the Preferred Embodiment

The practice of the present invention encompasses conventional techniques of chemistry, 30 immunology, molecular biology, biochemistry, protein chemistry, and recombinant DNA technology, which are within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Oligonucleotide Synthesis (M. Gait ed. 1984); Nucleic Acid Hybridization (B. Hames & S. Higgins, eds., 1984); Sambrook, Fritsch & Maniatis, Molecular Cloning: A Laboratory Manual, Second Edition (1989); PCR Technology (H.A. Erlich ed., Stockton Press); R. Scope, Protein Purification Principles and Practice (Springer-Verlag); and the series Methods in Enzymology (S. Colowick and N. Kaplan eds., Academic Press, Inc.).

10

Definitions

As used interchangeably herein, the terms "nucleic acid" "oligonucleotide", and "polynucleotides" include RNA, DNA, or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form. The term "nucleotide" as used herein as an adjective to 5 describe molecules comprising RNA, DNA, or RNA/DNA hybrid sequences of any length in singlestranded or duplex form. The term "nucleotide" is also used herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a phosphate group, or phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide. Although the term "nucleotide" is also used herein to encompass "modified nucleotides" which comprise at least one modifications (a) an alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar, for examples of analogous linking groups, purine, pyrimidines, and sugars see for example PCT publication No. WO 95/04064. However, the polynucleotides of the invention are preferably 15 comprised of greater than 50% conventional deoxyribose nucleotides, and most preferably greater than 90% conventional deoxyribose nucleotides. The polynucleotide sequences of the invention is prepared by any known method, including synthetic, recombinant, ex vivo generation, or a combination thereof, as well as utilizing any purification methods known in the art.

As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

The term "purified" is used herein to describe a polynucleotide or polynucleotide vector of the invention which has been separated from other compounds including, but not limited to other nucleic acids, charbohydrates, lipids and proteins (such as the enzymes used in the synthesis of the polynucleotide), or the separation of covalently closed polynucleotides from linear polynucleotides. A polynucleotide is substantially pure when at least about 50 %, preferably 60 to 75% of a sample exhibits a single polynucleotide sequence and conformation (linear versus covalently close). A substantially pure polynucleotide typically comprises about 50 %, preferably 60 to 90% weight/weight of a nucleic acid sample, more usually about 95%, and preferably is over about 99% pure. Polynucleotide purity or homogeneity is indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single polynucleotide band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art.

35 The term "polypeptide" refers to a polymer of amino without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide.

This term also does not specify or exclude prost-expression modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

As used herein, the term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring).

The term "purified" is used herein to describe a polypeptide of the invention which has been separated from other compounds including, but not limited to nucleic acids, lipids, charbohydates and other proteins. A polypeptide is substantially pure when at least about 50 %, preferably 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure polypeptide typically comprises about 50 %, preferably 60 to 90% weight/weight of a protein sample, more usually about 95%, and preferably is over about 99% pure. Polypeptide purity or homogeneity is indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art.

As used herein, the term "non-human animal" refers to any non-human vertebrate, birds and more usually mammals, preferably primates, farm animals such as swine, goats, sheep, donkeys, and horses, rabbits or rodents, more preferably rats or mice. As used herein, the term "animal" is used to refer to any vertebrate, preferable a mammal. Both the terms "animal" and "mammal" expressly embrace human subjects unless preceded with the term "non-human".

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where an antibody binding domain is formed from the folding of variable domains of an antibody molecule to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigenic determinant of an antigen., which allows an immunological reaction with the antigen. Antibodies 30 include recombinant proteins comprising the binding domains, as wells as fragments, including Fab, Fab', F(ab)2, and F(ab')2 fragments.

As used herein, an "antigenic determinant" is the portion of an antigen molecule, in this case an PG1 polypeptide, that determines the specificity of the antigen-antibody reaction. An "epitope" refers to an antigenic determinant of a polypeptide. An epitope can comprise as few as 3 amino acids 35 in a spatial conformation which is unique to the epitope. Generally an epitope consists of at least 6 such amino acids, and more usually at least 8-10 such amino acids. Methods for determining the

10

20

5

amino acids which make up an epitope include x-ray crystallography, 2-dimensional nuclear magnetic resonance, and epitope mapping e.g. the Pepscan method described by H. Mario Geysen et al. 1984. Proc. Natl. Acad. Sci. U.S.A. 81:3998-4002; PCT Publication No. WO 84/03564; and PCT Publication No. WO 84/03506.

The term "DNA construct" and "vector" are used herein to mean a purified or isolated polynucleotide that has been artificially designed and which comprises at least two nucleotide sequences that are not found as contiguous nucleotide sequences in their natural environment.

The terms "trait" and "phenotype" are used interchangeably herein and refer to any visible, detectable or otherwise measurable property of an organism such as symptoms of, or susceptibility to a disease for example. Typically the terms "trait" or "phenotype" are used herein to refer to symptoms of, or susceptibility to cancer or prostate cancer; or to refer to an individual's response to an anti-cancer agent or an anti-prostate cancer agent; or to refer to symptoms of, or susceptibility to side effects to an anticancer agent or an anti-prostate cancer agent.

The term "allele" is used herein to refer to variants of a nucleotide sequence. A biallelic polymorphism has two forms. Typically the first identified allele is designated as the original allele whereas other alleles are designated as alternative alleles. Diploid organisms is homozygous or heterozygous for an allelic form.

The term "heterozygosity rate" is used herein to refer to the incidence of individuals in a population, which are heterozygous at a particular allele. In a biallelic system the heterozygosity rate is on average equal to $2P_a(1-P_a)$, where P_a is the frequency of the least common allele. In order to be useful in genetic studies a genetic marker should have an adequate level of heterozygosity to allow a reasonable probability that a randomly selected person will be heterozygous.

The term "genotype" as used herein refers the identity of the alleles present in an individual or a sample. In the context of the present invention a genotype preferably refers to the description of the biallelic marker alleles present in an individual or a sample. The term "genotyping" a sample or an individual for a biallelic marker consists of determining the specific allele or the specific nucleotide carried by an individual at a biallelic marker.

The term "mutation" as used herein refers to a difference in DNA sequence between or among different genomes or individuals which has a frequency below 1%.

The term "haplotype" refers to a combination of alleles present in an individual or a sample. In the context of the present invention a haplotype preferably refers to a combination of biallelic marker alleles found in a given individual and which is associated with a phenotype.

The term "polymorphism" as used herein refers to the occurrence of two or more alternative genomic sequences or alleles between or among different genomes or individuals. "Polymorphic" refers to the condition in which two or m... variants of a specific genomic sequence can be found in a population. A "polymorphic site" is the locus at which the variation occurs. A single nucleotide

polymorphism is a single base pair change. Typically a single nucleotide polymorphism is the replacement of one nucleotide by another nucleotide at the polymorphic site. Deletion of a single nucleotide or insertion of a single nucleotide, also give rise to single nucleotide polymorphisms. In the context of the present invention "single nucleotide polymorphism" preferably refers to a single nucleotide substitution. Typically, between different genomes or between different individuals, the polymorphic site is occupied by two different nucleotides.

The terms "biallelic polymorphism" and "biallelic marker" are used interchangeably herein to refer to a nucleotide polymorphism having two alleles at a fairly high frequency in the population. A "biallelic marker allele" refers to the nucleotide variants present at a biallelic marker site. Usually a biallelic marker is a single nucleotide polymorphism. However, less commonly there are also insertions and deletions of up to 5 nucleotides which constitute biallelic markers for the purposes of the present invention. Typically the frequency of the less common allele of the biallelic markers of the present invention has been validated to be greater than 1%, preferably the frequency is greater than 10%, more preferably the frequency is at least 20% (i.e. heterozygosity rate of at least 0.32), even more preferably the frequency is at least 30% (i.e. heterozygosity rate of at least 0.42). A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker."

The location of nucleotides in a polynucleotide with respect to the center of the polynucleotide are described herein in the following manner. When a polynucleotide has an odd number of nucleotides, the nucleotide at an equal distance from the 3' and 5' ends of the polynucleotide is considered to be "at the center" of the polynucleotide, and any nucleotide immediately adjacent to the nucleotide at the center, or the nucleotide at the center itself is considered to be "within 1 nucleotide of the center." With an odd number of nucleotides in a polynucleotide any of the five nucleotides positions in the middle of the polynucleotide would be considered to be within 2 nucleotides of the center, and so on. When a polynucleotide has an even number of nucleotides, there would be a bond and not a nucleotide at the center of the polynucleotide. Thus, either of the two central nucleotides would be considered to be "within 1 nucleotide of the center" and any of the four nucleotides in the middle of the polynucleotide would be considered to be "within 2 nucleotides of the center", and so on.

The term "upstream" is used herein to refer to a location which is toward the 5' end of the polynucleotide from a specific reference point.

The terms "base paired" and "Watson & Crick base paired" are used interchangeably herein to refer to nucleotides which can be hydrogen bonded to one another be virtue of their sequence identities in a manner like that found in double-helical DNA with thymine or uracil residues linked to adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds (See Stryer, L., *Biochemistry*, 4th edition, 1995).

The terms "complementary" or "complement thereof" are used herein to refer to the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. This term is applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind.

As used herein the term "PG1-related biallelic marker" relates to a set of biallelic markers in linkage disequilibrium with PG1. The term PG1-related biallelic marker includes all of the biallelic markers used in the initial association studies shown below in Section I.D., including those biallelic markers contained in SEQ ID NOs: 21-38 and 57-62. The term PG1-related biallelic marker encompasses all of the following polymorphisms positioned in SEQ ID 179, and listed by internal reference number, including: 5-63-169 G or C in position 2159;

5-63-453 C or T in position 2443; 99-622-95 T or C in position 4452;

99-621-215 T or C in position 5733; 99-619-141 G or A in position 8438;

4-76-222 deletion of GT in position 11843; 4-76-361 C or T in position 11983;

15 4-77-151 G or C in position 12080; 4-77-294 A or G in position 12221;

4-71-33 G or T in position 12947;4-71-233 A or G in position 13147;

4-71-280 G or A in position 13194; 4-71-396 G or C in position 13310;

4-72-127 A or G in position 13342; 4-72-152 A or G in position 13367;

4-72-380 deletion of A in position 13594; 4-73-134 G or C in position 13680;

20 4-73-356 G or C in position 13902; 99-610-250 T or C in position 16231;

99-610-93 A or T in position 16388; 99-609-225 A or T in position 17608;

4-90-27 A or C in position 18034; 4-90-283 A or C in position 18290;

99-607-397 T or C in position 18786; 99-602-295 deletion of A in position 22835;

99-602-258 T or C in position 22872;

25 99-600-492 deletion of TATTG in position 25183;

99-600-483 T or G in position 25192; 5-23-288 A or G in position 25614;

99-598-130 T or C in position 26911; 99-592-139 A or T in position 32703;

99-217-277 C or T in position 34491; 5-47-284 A or G in position 34756;

99-589-267 T or G in position 34934; 99-589-41 G or C in position 35160;

30 99-12899-307 C or T in position 39897; 4-12-68 A or G in position 40598;

99-582-263 T or C in position 40816; 99-582-132 T or C in position 40947;

99-576-421 G or C in position 45783; 4-13-51 C or T in position 47929;

4-13-328 A or T in position 48206; 4-13-329 G or C in position 48207;

99-12903-381 C or T in position 49282; 5-56-208 A or G in position 50037;

35 5-56-225 A or G in position 50054; 5-56-272 A or G in position 50101;

5-56-391 G or T in position 50220; 4-61-269 A or G in position 50440;

4-61-391 A or G in position 50562; 4-63-99 A or G in position 50653;

4-62-120 A or G in position 50660; 4-62-205 A or G in position 50745;

4-64-113 A or T in position 50885; 4-65-104 A or G in position 51249;

5-28-300 A or G in position 51333; 5-50-269 C or T in position 51435;

5 4-65-324 C or T in position 51468; 5-71-129 G or C in position 51515;

5-50-391 G or C in position 51557; 5-71-180 A or G in position 51566;

4-67-40 C or T in position 51632; 5-71-280 A or C in position 51666;

5-58-167 A or G in position 52016; 5-30-325 C or T in position 52096;

5-58-302 A or T in position 52151; 5-31-178 A or G in position 52282;

10 5-31-244 A or G in position 52348; 5-31-306 deletion of A in position 52410;

5-32-190 C or T in position 52524; 5-32-246 C or T in position 52580;

5-32-378 deletion of A in position 52712; 5-53-266 G or C in position 52772;

5-60-158 C or T in position 52860; 5-60-390 A or G in position 53092;

5-68-272 G or C in position 53272; 5-68-385 A or T in position 53389;

15 5-66-53 deletion of GA in position 53511; 5-66-142 G or C in position 53600;

5-66-207 A or G in position 53665; 5-37-294 A or G in position 53815;

5-62-163 insertion of A in position 54365; 5-62-340 A or T in position 54541; and the compliments

thereof. The term PG1-related biallelic marker also includes all of the following biallelic markers

listed by internal reference number, and two SEQ ID NOs each of which contains a 47-mers with one

20 of the two alternative bases at position 24:

4-14-107 of SEQ ID NOs 185 and 262; 4-14-317 of SEQ ID NOs 186 and 263;

4-14-35 of

SEQ ID NOs 187 and 264; 4-20-149 of SEQ ID NOs 188 and 265;

4-20-77 of SEQ ID NOs 189and 266; 4-22-174 of SEQ ID NOs 190 and 267;

4-22-176 of SEQ ID NOs 191 and 268; 4-26-60 of SEQ ID NOs 192 and 269;

25 4-26-72 of SEQ ID NOs 193 and 270; 4-3-130 of SEQ ID NOs 194 and 271;

4-38-63 of SEQ ID NOs 195 and 272;

4-38-83 of SEQ ID NOs 196 and 273; 4-4-152 of SEQ ID NOs 197 and 274;

4-4-187 of SEQ ID NOs 198 and 275; 4-4-288 of SEQ ID NOs 199 and 276;

4-42-304 of SEQ ID NOs 200 and 277; 4-42-401 of SEQ ID NOs 201 and 278;

30 4-43-328 of SEQ ID NOs 202 and 279; 4-43-70 of SEQ ID NOs 203 and 280;

4-50-209 of SEQ ID NOs 204 and 281; 4-50-293 of SEQ ID NOs 205 and 282;

4-50-323 of SEQ ID NOs 206 and 283; 4-50-329 of SEQ ID NOs 207 and 284;

4-50-330 of SEQ ID NOs 208 and 285; 4-52-163 of SEQ ID NOs 209 and 286;

4-52-88 of SEQ ID NOs 210 and 287; 4-53-258 of SEQ ID NOs 211 and 288;

35 4-54-283 of SEQ ID NOs 212 and 289; 4-54-388 of SEQ ID NOs 213 and 290;

4-55-70 of SEQ ID NOs 214 and 291; 4-55-95 of SEQ ID NOs 215 and 292;

22

4-56-159 of SEQ ID NOs 216 and 293; 4-56-213 of SEQ ID NOs 217 and 294;

4-58-289 of SEQ ID NOs 218 and 295; 4-58-318 of SEQ ID NOs 219 and 296;

4-60-266 of SEQ ID NOs 220 and 297; 4-60-293 of SEQ ID NOs 221 and 298;

4-84-241 of SEQ ID NOs 222 and 299; 4-84-262 of SEQ ID NOs 223 and 300;

5 4-86-206 of SEQ ID NOs 224 and 301; 4-86-309 of SEQ ID NOs 225 and 302;

4-88-349 of SEQ ID NOs 226 and 303; 4-89-87 of SEQ ID NOs 227 and 304;

99-123-184 of SEQ ID NOs 228 and 305; 99-128-202 of SEQ ID NOs 229 and 306;

99-128-275 of SEQ ID NOs 230 and 307; 99-128-313 of SEQ ID NOs 231 and 308; 99-128-60 of

SEQ ID NOs 232 and 309; 99-12907-295 of SEQ ID NOs 233 and 310; 99-130-58 of SEQ ID NOs

10 234 and 311; 99-134-362 of SEQ ID NOs 235 and 312;

99-140-130 of SEQ ID NOs 236 and 313; 99-1462-238 of SEQ ID NOs 237 and 314; 99-147-181 of SEQ ID NOs 238 and 315; 99-1474-156 of SEQ ID NOs 239 and 316; 99-1474-359 of SEQ ID NOs 240 and 317; 99-1479-158 of SEQ ID NOs 241 and 318; 99-1479-379 of SEQ ID NOs 242 and 319; 99-148-129 of SEQ ID NOs 243 and 320; 99-148-132 of SEQ ID NOs 244 and 321; 99-148-139

15 of SEQ ID NOs 245 and 322;

99-148-140 of SEQ ID NOs 246 and 323; 99-148-182 of SEQ ID NOs 247 and 324;

99-148-366 of SEQ ID NOs 248 and 325; 99-148-76 of SEQ ID NOs 249 and 326;

99-1480-290 of SEQ ID NOs 250 and 327; 99-1481-285 of SEQ ID NOs 251 and 328; 99-1484-101 of SEQ ID NOs 252 and 329; 99-1484-328 of SEQ ID NOs 253 and 330; 99-1485-251 of SEQ ID NOs

20 254 and 331; 99-1490-381 of SEQ ID NOs 255 and 332; 99-1493-280 of SEQ ID NOs 256 and 333; 99-151-94 of SEQ ID NOs 257 and 334;

99-211-291 of SEQ ID NOs 258 and 335; 99-213-37 of SEQ ID NOs 259 and 336;

99-221-442 of SEQ ID NOs 260 and 337; 99-222-109 of SEQ ID NOs 261 and 338; and the compliments thereof.

The term "non-genic" is used herein to describe PG1-related biallelic markers, as well as polynucleotides and primers which do not occur in the human PG1 genomic sequence of SEQ ID NO: 179. The term "genic" is used herein to describe PG1-related biallelic markers as well as polynucleotides and primers which do occur in the human PG1 genomic sequence of SEQ ID NO: 179.

The terms "an anti-cancer agent" refers to a drug or a compound that is capable of reducing the growth rate, rate of metastasis, or viability of tumor cells in a mammal, is capable of reducing the size or eliminating tumors in a mammal, or is capable of increasing the average life span of a mammal or human with cancer. Anti-cancer agents also include compounds which are able to reduce the risk of cancer developing in a population, particularly a high risk population. The terms "an anti-prostate cancer agent" is an anti-cancer agent that has these effects on cells or tumors that are derived from prostate cancer cells.

5

The terms "response to an anti-cancer agent" and "response to an anti- prostate cancer agent" refer to drug efficacy, including but not limited to ability to metabolize a compound, to the ability to convert a pro-drug to an active drug, and to the pharmacokinetics (absorption, distribution, elimination) and the pharmacodynamics (receptor-related) of a drug in an individual.

The terms "side effects to an anti-cancer agent" and "side effects to an anti-prostate cancer agent" refer to adverse effects of therapy resulting from extensions of the principal pharmacological action of the drug or to idiosyncratic adverse reactions resulting from an interaction of the drug with unique host factors. These side effects include, but are not limited to, adverse reactions such as dermatological, hematological or hepatological toxicities and further includes gastric and intestinal 10 ulceration, disturbance in platelet function, renal injury, nephritis, vasomotor rhinitis with profuse watery secretions, angioneurotic edema, generalized urticaria, and bronchial asthma to laryngeal edema and bronchoconstriction, hypotension, sexual dysfunction, and shock.

As used herein the term "homology" refers to comparisons between protein and/or nucleic acid sequences and is evaluated using any of the variety of sequence comparison algorithms and 15 programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, 1988, Proc. Natl. Acad. Sci. USA 85(8):2444-2448; Altschul et al., 1990, J. Mol. Biol. 215(3):403-410; Thompson et al., 1994, Nucleic Acids Res. 22(2):4673-4680; Higgins et al., 1996, Methods Enzymol. 266:383-402; Altschul et al., 1990, J. Mol. Biol. 215(3):403-410; Altschul et al., 1993, Nature Genetics 3:266-272). In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") which is well known in the art (see, e.g., Karlin and Altschul, 1990, Proc. Natl. Acad. Sci. USA 87:2267-2268; Altschul et al., 1990, J. Mol. Biol. 215:403-410; Altschul et al., 1993, Nature Genetics 3:266-272; Altschul et al., 1997, Nuc. Acids Res. 25:3389-3402). In particular, five specific BLAST programs are used to perform the following 25 task:

- (1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;
- (2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;
- (3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database:
- (4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

(5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (i.e., aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet et al., 1992, Science 256:1443-1445; Henikoff and Henikoff, 1993, Proteins 17:49-61). Less preferably, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978, Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington: National Biomedical Research Foundation). The BLAST programs evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, e.g., Karlin and Altschul, 1990, Proc. Natl. Acad. Sci. USA 87:2267-2268).

I. ISOLATION AND CHARACTERIZATION OF THE PG1 GENE AND PROTEINS I.A. The 8p23 Region- LOH Studies: Implications of 8p23 Region in Distinct Cancer Types

Substantial amounts of LOH data support the hypothesis that genes associated with distinct 20 cancer types are located within 8p23 region of the human genome. Emi et al., demonstrated the implication of 8p23.1-8p21.3 region in cases of hepatocellular carcinoma, colorectal cancer, and nonsmall cell lung cancer. (Emi M, Fujiwara Y, Nakajima T, Tsuchiya E, Tsuda H, Hirohashi S, Maeda Y, Tsuruta K, Miyaki M, Nakamura Y, Cancer Res. 1992 Oct 1; 52(19): 5368-5372) Yaremko, et al., 25 showed the existence of two major regions of LOH for chromosome 8 markers in a sample of 87 colorectal carcinomas. The most prominent loss was found for 8p23.1-pter, where 45% of informative cases demonstrated loss of alleles. (Yaremko ML, Wasylyshyn ML, Paulus KL, Michelassi F, Westbrook CA, Genes Chromosomes Cancer 1994 May;10(1):1-6). Scholnick et al. demonstrated the existence of three distinct regions of LOH for the markers of chromosome 8 in cases of squamous cell carcinoma of the supraglottic larynx. They showed that the allelic loss of 8p23 marker D8S264 serves as a statistically significant, independent predictor of poor prognosis for patients with supraglottic squamous cell carcinoma. (Scholnick SB, Haughey BH, Sunwoo JB, el-Mofty SK, Baty JD, Piccirillo JF, Zequeira MR, J. Natl. Cancer Inst. 1996 Nov 20; 88(22): 1676-1682 and Sunwoo JB, Holt MS, Radford DM, Deeker C, Scholnick S, Genes Chromosomes Cancer 1996

35 Jul; 16(3):164-169).

In other studies, Nagai et al. demonstrated the highest loss of heterozygosity in the specific region of 8p23 by genome wide scanning of LOH in 120 cases of hepatocellular carcinoma (HCC). (Nagai H, Pineau P, Tiollais P, Buendia MA, Dejean A, Oncogene 1997 Jun 19; 14(24): 2927-2933). Gronwald et al. demonstrated 8p23-pter loss in renal clear cell carcinomas. (Gronwald J, Storkel S, Holtgreve-Grez H, Hadaczek P, Brinkschmidt C, Jauch A, Lubinski J, Cremer, Cancer Res. 1997 Feb 1; 57(3): 481-487).

The same region is involved in specific cases of prostate cancer. Matsuyama et al. showed the specific deletion of the 8p23 band in prostate cancer cases, as monitored by FISH with D8S7 probe. (Matsuyama H, Pan Y, Skoog L, Tribukait B, Naito K, Ekman P, Lichter P, Bergerheim US Oncogene 1994 Oct; 9(10): 3071-3076). They were able to document a substantial number of cases with deletions of 8p23 but retention of the 8p22 marker LPL. Moreover, Ichikawa et al. deduced the existence of a prostate cancer metastasis suppressor gene and localized it to 8p23-q12 by studies of metastasis suppression in highly metastatic rat prostate cells after transfer of human chromosomes. (Ichikawa T. Nihei N, Kuramochi H, Kawana Y, Killary AM, Rinker-Schaeffer CW, Barrett JC, Isaacs JT, Kugoh H, Oshimura M, Shimazaki J, Prostate Suppl. 1996; 6: 31-35).

Recently Washbum et al. were able to find substantial numbers of tumors with the allelic loss specific to 8p23 by LOH studies of 31 cases of human prostate cancer. (Washbum J, Woino K, and Macoska J, Proceedings of American Association for Cancer Research, March 1997; 38). In these samples they were able to define the minimal overlapping region with deletions covering genetic interval D8S262-D8S277.

Linkage Analysis Studies: Search for Prostate Cancer

Linked Regions on Chromosome 8

Microsatellite markers mapping to chromosome 8 were used by the inventors to perform linkage analysis studies on 194 individuals issued from 47 families affected with prostate cancer. While multiple point analysis led to weak linkage results, two point lod score analysis led to non significant results, as shown below.

Two point lod (parametric analysis)

MARKER Di	stance (cM) Z(lod) scores
D8S1742	-0.13
D8S56# . 0.6	-0.07
# of families analyzed	47
Total # of individuals	194
genotyped	
Total # of affected individuals gene	otyped 122

In view of the non-significant results obtained with linkage analysis, a new mapping approach based on linkage disequilibrium of biallelic markers was utilised to identify genes responsible for sporadic cases of prostate cancer.

WO 99/32644

I.B. Linkage Disequilibrium Using Biallelic Markers To Identify Candidate Loci Responsible 5 For Disease

Linkage Disequilibrium

Once a chromosomal region has been identified as potentially harboring a candidate gene associated with a sporadic trait, an excellent approach to refine the candidate gene's location within the identified region is to look for statistical associations between the trait and some marker genotype when comparing an affected (trait +) and a control (trait -) population.

Association studies have most usually relied on the use of biallelic markers. Biallelic markers are genome-derived polynucleotides that exhibit biallelic polymorphism at one single base position. By definition, the lowest allele frequency of a biallelic polymorphism is 1%; sequence variants that show allele frequencies below 1% are called rare mutations. There are potentially more than 10⁷ biallelic markers lying along the human genome.

Association studies seek to establish correlations between traits and genetic markers and are based on the phenomenon of linkage disequilibrium (LD). LD is defined as the trend for alleles at nearby loci on haploid genomes to correlate in the population. If two genetic loci lie on the same chromosome, then sets of alleles on the same chromosomal segment (i.e., haplotypes) tend to be transmitted as a block from generation to generation. When not broken up by recombination, haplotypes can be tracked not only through pedigrees but also through populations. The resulting phenomenon at the population level is that the occurrence of pairs of specific alleles at different loci on the same chromosome is not random, and the deviation from random is called linkage disequilibrium.

Since results generated by association studies are essentially based on the quantitative calculation of allele frequencies, they best apply to the analysis of germline mutations. This is mainly due to the fact that allelic frequencies are difficult to quantify within tumor tissue samples because of the usual presence of normal cells within the studied tumor samples. Association studies applied to cancer genetics will therefore be best suited to the identification of tumor suppressor genes.

Trait Localization by Linkage Disequilibrium Mapping

Any gene responsible or partly responsible for a given trait will be in LD with some flanking markers. To map such a gene, specific alleles of these flanking markers which are associated with the gene or genes responsible for the trait are identified. Although the following discussion of techniques for finding the gene or genes associated with a particular trait using linkage disequilibrium mapping, refers to locating a single gene which is responsible for the trait, it will be appreciated that the same techniques may also be used to identify genes which are partially responsible for the trait.

25

PCT/IB98/02133 WO 99/32644 27

Association studies is conducted within the general population (as opposed to the linkage analysis techniques discussed above which are limited to studies performed on related individuals in one or several affected families).

Association between a biallelic marker A and a trait T may primarily occur as a result of three 5 possible relationships between the biallelic marker and the trait. First, allele a of biallelic marker A is directly responsible for trait T (e.g., Apo E e4 allele and Alzheimer's disease). However, since the majority of the biallelic markers used in genetic mapping studies are selected randomly, they mainly map outside of genes. Thus, the likelihood of allele a being a functional mutation directly related to trait T is therefore very low.

An association between a biallelic marker A and a trait T may also occur when the biallelic marker is very closely linked to the trait locus. In other words, an association occurs when allele a is in linkage disequilibrium with the trait-causing allele. When the biallelic marker is in close proximity to a gene responsible for the trait, more extensive genetic mapping will ultimately allow a gene to be discovered near the marker locus which carries mutations in people with trait T (i.e. the gene 15 responsible for the trait or one of the genes responsible for the trait). As will be further exemplified below using a group of biallelic markers which are in close proximity to the gene responsible for the trait, the location of the causal gene can be deduced from the profile of the association curve between the biallelic markers and the trait. The causal gene will be found in the vicinity of the marker showing the highest association with the trait.

Finally, an association between a biallelic marker and a trait may occur when people with the trait and people without the trait correspond to genetically different subsets of the population who, coincidentally, also differ in the frequency of allele a (population stratification). This phenomenon is avoided by using large heterogeneous samples.

Association studies are particularly suited to the efficient identification of susceptibility genes that present common polymorphisms, and are involved in multifactorial traits whose frequency is 25 relatively higher than that of diseases with monofactorial inheritance.

Application of Linkage Disequilibrium Mapping

to Candidate Gene Identification

The general strategy of association studies using a set of biallelic markers, is to scan two pools of individuals (affected individuals and unaffected controls) characterized by a well defined phenotype in order to measure the allele frequencies for a number of the chosen markers in each of these pools. If a positive association with a trait is identified using an array of biallelic markers having a high enough density, the causal gene will be physically located in the vicinity of the associated markers, since the markers showing positive association to the trait are in linkage disequilibrium with the trait locus. Regions harboring a gene responsible for a particular trait which

10

are identified through association studies using high density sets of biallelic markers will, on average, be 20 - 40 times shorter in length than those identified by linkage analysis.

Once a positive association is confirmed as described above, BACs (bacterial artificial chromosomes) obtained from human genomic libraries, constructed as described below, harboring the markers identified in the association analysis are completely sequenced.

Once a candidate region has been sequenced and analyzed, the functional sequences within the candidate region (exons and promoters, and other potential regulatory regions) are scanned for mutations which are responsible for the trait by comparing the sequences of a selected number of controls and affected individuals using appropriate software. Candidate mutations are further confirmed by screening a larger number of affected individuals and controls using the microsequencing techniques described below.

Candidate mutations are identified as follows. A pair of oligonucleotide primers is designed in order to amplify the sequences of every predicted functional region. PCR amplification of each predicted functional sequence is carried out on genomic DNA samples from affected patients and unaffected controls. Amplification products from genomic PCR are subjected to automated dideoxy terminator sequencing reactions and electrophoresed on ABI 377 sequencers. Following gel image analysis and DNA sequence extraction, the sequence data are automatically analyzed to detect the presence of sequence variations among affected cases and unaffected controls. Sequences are systematically verified by comparing the sequences of both DNA strands of each individual.

Polymorphisms are then verified by screening a larger population of affected individuals and controls using the microsequencing technique described below in an individual test format. Polymorphisms are considered as candidate mutations when present in affected individuals and controls at frequencies compatible with the expected association results.

Association Studies: Statistical Analysis and Haplotyping

As mentioned above, linkage analysis typically localizes a disease gene to a chromosomal region of several megabases. Further refinement in location requires the analysis of additional families in order to increase the number of recombinants. However, this approach becomes unfeasible because recombination is rarely observed even within large pedigrees (Boehnke, M, 1994, Am. J. Hum. Genet. 55: 379-390).

Linkage disequilibrium, the nonrandom association of alleles at linked loci, may offer an alternative method of obtaining additional recombinants. When a chromosome carrying a mutant allele of a gene responsible for a given trait is first introduced into a population as a result of either mutation or migration, the mutant allele necessarily resides on a chromosome having a unique set of linked markers (haplotype). Consequently, there is complete disequilibrium between these markers and the disease mutation: the disease mutation is present only linked to a specific set of marker alleles. Through subsequent generations, recombinations occur between the disease mutation and

20

29

these marker polymorphisms, resulting in a gradual disappearance of disequilibrium. The degree of disequilibrium dissipation depends on the recombination frequency, so the markers closest to the disease gene will tend to show higher levels of disequilibrium than those that are farther away (Jorde LB, 1995, Am. J. Hum. Genet. 56: 11-14). Because linkage disequilibrium patterns in a present-day population reflect the action of recombination through many past generations, disequilibrium analysis effectively increases the sample of recombinants. Thus the mapping resolution achieved through the analysis of linkage disequilibrium patterns is much higher than that of linkage analysis.

In practice, in order to define the regions bearing a candidate gene, the affected and control populations are genotyped using an appropriate number of biallelic markers (at a density of 1 marker every 50-150 kilobases). Then, a marker/trait association study is performed that compares the genotype frequency of each biallelic marker in the affected and control populations by means of a chi square statistical test (one degree of freedom).

After the first screening, additional markers within the region showing positive association are genotyped in the affected and control populations. Two types of complementary analysis are then performed. First, a marker/trait association study (as described above) is performed to refine the location of the gene responsible for the trait. In addition, a haplotype association analysis is performed to define the frequency and the type of the ancestral/preferential carrier haplotype. Haplotype analysis, by combining the informativeness of a set of biallelic markers increases the power of the association analysis, allowing false positive and/or negative data that may result from the single marker studies to be eliminated.

The haplotype analysis is performed by estimating the frequencies of all possible haplotypes for a given set of biallelic markers in the case and control populations, and comparing these frequencies by means of a chi square statistical test (one degree of freedom). Haplotype estimations are performed by applying the Expectation-Maximization (EM) algorithm (Excoffier L & Slatkin M, 1995, Mol. Biol. Evol. 12: 921-927), using the EM-HAPLO program (Hawley ME, Pakstis AJ & Kidd KK, 1994, Am. J. Phys. Anthropol. 18: 104). The EM algorithm is used to estimate haplotype frequencies in the case when only genotype data from unrelated individuals are available. The EM algorithm is a generalized iterative maximum likelihood approach to estimation that is useful when data are ambiguous and/or incomplete.

30 The application of biallelic marker based linkage disequilibrium analysis to the 8p23 region to identify a gene associated with prostate cancer is described below.

I.C. Application of Linkage Disequilibrium Mapping to the 8p23 Region

YAC Contig Construction in 8p23 Region

First, a YAC contig which contains the 8p23 region was constructed as follows. The CEPH-35 Jenethon YAC map for the entire human genome (Chumakov I.M. et al. A YAC contig map of the human genome, Nature, 377 Supp.: 175-297, 1995) was used for detailed contig building in the region

around D8S262 and D8S277 genetic markers. Screening data available for regional genetic markers D8S1706, D8S277, D8S1742, D8S518, D8S262, D8S1798, D8S1140, D8S561 and D8S1819 were used to select the following set of CEPH YACs, localized within this region: 832_g_12, 787_c_11, 920_h_7, 807_a_1, 842_b_1, 745_a_3, 910_d_3, 879_f_11, 918_c_6, 764_c_7, 910_f_12, 967_c_11, 5 856_d_8, 792_a_6, 812_h_4, 873_c_8, 930_a_2, 807_a_1, 852_d_10. This set of YACs was tested by PCR with the above mentioned genetic markers as well as with other publicly available markers supposedly located within the 8p23 region. As a result of these studies, a YAC STS contig map was generated around genetic markers D8S262 and D8S277. The two CEPH YACs, 920_h_7 (1170 kb insert size) and 910_f_12 (1480 kb insert size) constitute a minimal tiling path in this region, with an 10 estimated size of ca. 2 Megabases.

During this mapping effort, the following publicly known STS markers were precisely located within the contig: WI-14718, WI-3831, D8S1413E, WI-8327, WI-3823, ND4.

BAC Contig Construction Covering D8S262-D8S277

Fragment Within 8p23 Region of the Human Genome

Following construction of the YAC contig, a BAC contig was constructed as follows. BAC libraries were obtained as described in Woo et al. Nucleic Acids Res., 1994, 22, 4922-4931. Briefly, two different whole human genome libraries were produced by cloning BamHI or HindIII partially digested DNA from a lymphoblastoid cell line (derived from individual N°8445, CEPH families) into the pBeloBAC11 vector (Kim et al. Genomics, 1996, 34, 213-218). The library produced with the 20 BamHI partial digestion contained 110,000 clones with an average insert size of 150 kb, which corresponds to 5 human haploid genome equivalents. The library prepared with the HindIII partial digestion corresponds to 3 human genome equivalents with an average insert size of 150 kb.

BAC Screening

The human genomic BAC libraries obtained as described above were screened with all of the above mentioned STSs. DNA from the clones in both libraries was isolated and pooled in a three dimensional format ready for PCR screening with the above mentioned STSs using high throughput PCR methods (Chumakov et al., Nature 1995, 377: 175-298). Briefly, three dimensional pooling consists in rearranging the samples to be tested in a manner which allows the number of PCR reactions required to screen the clones with STSs to be reduced by at least 100 fold, as compared to screening each clone individually. PCR amplification products were detected by conventional agarose gel electrophoresis combined with automated image capturing and processing.

In a final step, STS-positive clones were checked individually. Subchromosomal localization of BACs was systematically verified by fluorescence in situ hybridization (FISH), performed on metaphasic chromosomes as described by Cherif et al. Proc. Natl. Acad. Sci. USA 1990, 87: 6639-6643.

35

BAC insert size was determined by Pulsed Field Gel Electrophoresis after digestion with restriction enzyme NotI.

BAC Contig Analysis

The ordered BACs selected by STS screening and verified by FISH, were assembled into contigs and new markers were generated by partial sequencing of insert ends from some of them. These markers were used to fill the gaps in the contig of BAC clones covering the chromosomal region around D8S277, having an estimated size of 2 megabases. Selected BAC clones from the contig were subcloned and sequenced.

BAC Subcloning

Each BAC human DNA was first extracted using the alkaline lysis procedure and then sheared by sonication. The obtained DNA fragments were end-repaired and electrophoresed on preparative agarose gels. The fragments in the desired size range were isolated from the gel, purified and ligated to a linearized, dephosphorylated, blunt-ended plasmid cloning vector (pBluescript II Sk (+)). Example 1 describes the BAC subcloning procedure.

15 Example 1

The cells obtained from three liters overnight culture of each BAC clone were treated by alkaline lysis using conventional techniques to obtain the BAC DNA containing the genomic DNA inserts. After centrifugation of the BAC DNA in a cesium chloride gradient, ca. 50µg of BAC DNA was purified. 5-10µg of BAC DNA was sonicated using three distinct conditions, to obtain fragments 20 of the desired size. The fragments were treated in a 50 μl volume with two units of Vent polymerase for 20 min at 70°C, in the presence of the four deoxytriphosphates (100µM). The resulting bluntended fragments were separated by electrophoresis on low-melting point 1% agarose gels (60 Volts for 3 hours). The fragments were excised from the gel and treated with agarase. After chloroform extraction and dialysis on Microcon 100 columns, DNA in solution was adjusted to a 100 ng/µl 25 concentration. A ligation was performed overnight by adding 100 ng of BAC fragmented DNA to 20 ng of pBluescript II Sk (+) vector DNA linearized by enzymatic digestion, and treated by alkaline phosphatase. The ligation reaction was performed in a 10 µl final volume in the presence of 40 units/µl T4 DNA ligase (Epicentre). The ligated products were electroporated into the appropriate cells (ElectroMAX E.coli DH10B cells). IPTG and X-gal were added to the cell mixture, which was 30 then spread on the surface of an ampicillin-containing agar plate. After overnight incubation at 37°C, recombinant (white) colonies were randomly picked and arrayed in 96 well microplates for storage and sequencing.

Partial Sequencing of BACs

At least 30 of the obtained BAC clones were sequenced by the end pair-wise method (500 bp sequence from each end) using a dye-primer cycle sequencing procedure. Pair-wise sequencing was performed until a map allowing the relative positioning of selected markers along the corresponding

DNA region was established. Example 2 describes the sequencing and ordering of the BAC inserts.

Example 2

The subclone inserts were amplified by PCR on overnight bacterial cultures, using vector primers flanking the insertions. The insert extremity sequences (on average 500 bases at each end) 5 were determined by fluorescent automated sequencing on ABI 377 sequencers, with a ABI Prism DNA Sequencing Analysis software (2.1.2 version).

The sequence fragments from BAC subclones were assembled using Gap4 software from R. Staden (Bonfield et al. 1995). This software allows the reconstruction of a single sequence from sequence fragments. The sequence deduced from the alignment of different fragments is called the 10 consensus sequence. We used directed sequencing techniques (primer walking) to complete sequences and link contigs.

Figure 1 shows the overlapping BAC subclones (labeled BAC) which make up the assembled contig and the positions of the publicly known STS markers along the contig.

Identification of Biallelic Markers Lying Along the BAC Contig

Following assembly of the BAC contig, biallelic markers lying along the contig were then identified. Given that the assessed distribution of informative biallelic markers in the human genome (biallelic polymorphisms with a heterozygosity rate higher than 42%) is one in 2.5 to 3 kb, six 500 bp genomic fragments have to be screened in order to identify 1 biallelic marker. Six pairs of primers per potential marker, each one defining a ca. 500 bp amplification fragment, were derived from the 20 above mentioned BAC partial sequences. All primers contained a common upstream oligonucleotide tail enabling the easy systematic sequencing of the resulting amplification fragments. Amplification of each BAC-derived sequence was carried out on pools of DNA from ca. 100 individuals. The conditions used for the polymerase chain reaction were optimized so as to obtain more than 95% of PCR products giving 500bp-sequence reads.

The amplification products from genomic PCR using the oligonucleotides derived from the BAC subclones were subjected to automated dideoxy terminator sequencing reactions using a dyeprimer cycle sequencing protocol. Following gel image analysis and DNA sequence extraction, sequence data were automatically processed with appropriate software to assess sequence quality and to detect the presence of biallelic sites among the pooled amplified fragments. Biallelic sites were systematically verified by comparing the sequences of both strands of each pool.

The detection limit for the frequency of biallelic polymorphisms detected by sequencing pools of 100 individuals is 0.3 +/- 0.05 for the minor allele, as verified by sequencing pools of known allelic frequencies. Thus, the biallelic markers selected by this method will be "informative biallelic markers" since they have a frequency of 0.3 to 0.5 for the minor allele and 0.5 to 0.7 for the major allele, therefore an average heterozygosity rate higher than 42%.

Example 3 describes the preparation of genomic DNA samples from the individuals screened

15

WO 99/32644

to identify biallelic markers.

5

Example 3

33

The population used in order to generate biallelic markers in the region of interest consisted of ca. 100 unrelated individuals corresponding to a French heterogeneous population.

DNA was extracted from peripheral venous blood of each donor as follows.

30 ml of blood were taken in the presence of EDTA. Cells (pellet) were collected after centrifugation for 10 minutes at 2000 rpm. Red cells were lysed by a lysis solution (50 ml final volume: 10 mM Tris pH7.6; 5 mM MgCl₂; 10 mM NaCl). The solution was centrifuged (10 minutes, 2000 rpm) as many times as necessary to eliminate the residual red cells present in the supernatant, after resuspension of the pellet in the lysis solution.

The pellet of white cells was lysed overnight at 42°C with 3.7 ml of lysis solution composed of:

- 3 ml TE 10-2 (Tris-HCl 10 mM, EDTA 2 mM) / NaCl 0.4 M
- 200 µl SDS 10%
- 15 500 μl K-proteinase (2 mg K-proteinase in TE 10-2 / NaCl 0.4 M).

For the extraction of proteins, 1 ml saturated NaCl (6M) (1/3.5 v/v) was added. After vigorous agitation, the solution was centrifuged for 20 minutes at 10000 rpm.

For the precipitation of DNA, 2 to 3 volumes of 100% ethanol were added to the previous supernatant, and the solution was centrifuged for 30 minutes at 2000 rpm. The DNA solution was rinsed three times with 70% ethanol to eliminate salts, and centrifuged for 20 minutes at 2000 rpm. The pellet was dried at 37°C, and resuspended in 1 ml TE 10-1 or 1 ml water. The DNA concentration was evaluated by measuring the OD at 260 nm (1 unit OD = $50 \mu g/ml$ DNA).

To determine the presence of proteins in the DNA solution, the OD 260 / OD 280 ratio was determined. Only DNA preparations having a OD 260 / OD 280 ratio between 1.8 and 2 were used in the subsequent steps described below.

DNA Amplification

Once each BAC was isolated, pairs of primers, each one defining a 500 bp-amplification fragment, were designed. Each of the primers contained a common oligonucleotide tail upstream of the specific bases targeted for amplification, allowing the amplification products from each set of primers to be sequenced using the common sequence as a sequencing primer. The primers used for the genomic amplification of sequences derived from BACs were defined with the OSP software (Hillier L. and Green P. Methods Appl., 1991, 1: 124-8). The synthesis of primers was performed following the phosphoramidite method, on a GENSET UFPS 24.1 synthesizer.

Example 4 provides the procedures used in the amplification reactions.

35 Example 4

The amplification of each sequence was performed by PCR (Polymerase Chain Reaction) as follows:

The amplification was performed on a Perkin Elmer 9600 Thermocycler or MJ Research PTC200 with heating lid. After heating at 94°C for 10 minutes, 35 cycles were performed. Each cycle comprised: 30 sec at 94°C, 1 minute at 55°C, and 30 sec at 72°C. For final elongation, 7 minutes at 72°C ended the amplification.

The obtained quantity of amplification products was determined on 96-well microtiter plates, using a fluorimeter and Picogreen as intercalating agent (Molecular Probes).

The sequences of the amplification products were determined for each of the approximately 100 individuals from whom genomic DNA was obtained. Those amplification products which contained biallelic markers were identified.

Figure 1 shows the locations of the biallelic markers along the 8p23 BAC contig. This first set of markers corresponds to a medium density map of the candidate locus, with an inter-marker distance averaging 50kb-150kb.

A second set of biallelic markers was then generated as described above in order to provide a very high-density map of the region identified using the first set of markers which can be used to conduct association studies, as explained below. The high density map has markers spaced on average every 2-50kb.

The biallelic markers were then used in association studies as described below.

Collection of DNA samples from affected and non-affected individuals

Prostate cancer patients were recruited according to clinical inclusion criteria based on pathological or radical prostatectomy records. Control cases included in this study were both ethnically- and age-matched to the affected cases; they were checked for both the absence of all clinical and biological criteria defining the presence or the risk of prostate cancer, and for the absence of related familial prostate cancer cases. Both affected and control individuals corresponded to unrelated cases.

The two following pools of independent individuals were used in the association studies. The first pool, comprising individuals suffering from prostate cancer, contained 185 individuals. Of these

20

185 cases of prostate cancer, 45 cases were sporadic and 140 cases were familial. The second pool, the control pool, contained 104 non-diseased individuals.

Haplotype analysis was conducted using additional diseased (total samples: 281) and control samples (total samples: 130), from individuals recruited according to similar criteria.

Genotyping Affected and Control Individuals

The general strategy to perform the association studies was to individually scan the DNA samples from all individuals in each of the two populations described above in order to establish the allele frequencies of the above described biallelic markers in each of these populations.

Allelic frequencies of the above-described biallelic markers in each population were determined by performing microsequencing reactions on amplified fragments obtained by genomic PCR performed on the DNA samples from each individual.

DNA samples and amplification products from genomic PCR were obtained in similar conditions as those described above for the generation of biallelic markers, and subjected to automated microsequencing reactions using fluorescent ddNTPs (specific fluorescence for each ddNTP) and the appropriate oligonucleotide microsequencing primers which hybridized just upstream of the polymorphic base. Once specifically extended at the 3' end by a DNA polymerase using the complementary fluorescent dideoxynucleotide analog (thermal cycling), the primer was precipitated to remove the unincorporated fluorescent ddNTPs. The reaction products were analyzed by electrophoresis on ABI 377 sequencing machines.

Example 5 describes one microsequencing procedure.

Example 5

5 μl of PCR products in a microtiter plate were added to 5 μl purification mix {2U SAP (Amersham); 2U Exonuclease I (Amersham); 1 μl SAP10X buffer: 400mM Tris-HCl pH8, 100 mM MgCl2; H2O final volume 5 μl}. The reaction mixture was incubated 30 minutes at 37°C, and denatured 10 minutes at 94°C. After 10 sec centrifugation, the microsequencing reaction was performed on line with the whole purified reaction mixture (10 μl) in the microplate using 10 pmol microsequencing oligonucleotide (23mers, GENSET, crude synthesis, 5 OD), 0.5 U Thermosequenase (Amersham), 1.25 μl Thermosequenase 16X buffer (Amersham), both of the fluorescent ddNTPs (Perkin Elmer) corresponding to the polymorphism {0.025 μl ddTTP and ddCTP, 0.05 μl ddATP and ddGTP}, H2O to a final volume of 20 μl. A PCR program on a GeneAmp 9600 thermocycler was carried out as follows: 4 minutes at 94°C; 5 sec at 55°C / 10 sec at 94°C for 20 cycles. The reaction product was incubated at 4°C until precipitation. The microtiter plate was centrifuged 10 sec at 1500 rpm. 19 μl MgCl2 2mM and 55 μl 100 % ethanol were added in each well. After 15 minute incubation at room temperature, the microtiter plate was centrifuged at 3300 rpm 15 minutes at 4°C. Supernatants were discarded by inverting the microtitre plate on a box folded to proper size and by centrifugation at 300 rpm 2 minutes at 4°C afterwards. The microplate was then dried 5 minutes in a

PCT/IB98/02133 36

vacuum drier. The pellets were resuspended in 2.5 μl formamide EDTA loading buffer (0.7 μl of 9 μ g/ μ l dextran blue in 25 mM EDTA and 1.8 μ l formamide). A 10% polyacrylamide gel / 12 cm / 64 wells was pre-run for 5 minutes on a 377 ABI 377 sequencer. After 5 minutes denaturation at 100°C, 0.8 µl of each microsequencing reaction product was loaded in each well of the gel. After migration 5 (2 h 30 for 2 microtiter plates of PCR products per gel), the fluorescent signals emitted by the incorporated ddNTPs were analyzed on the ABI 377 sequencer using the GENESCAN software (Perkin Elmer). Following gel analysis, data were automatically processed with a software that allowed the determination of the alleles of biallelic markers present in each amplified fragment.

I.D. Initial Association Studies

10

Association studies were run in two successive steps. In a first step, a rough localization of the candidate gene was achieved by determining the frequencies of the biallelic markers of Figure 1 in the affected and unaffected populations. The results of this rough localization are shown in Figure 2. This analysis indicated that a gene responsible for prostate cancer was located near the biallelic marker designated 4-67.

In a second phase of the analysis, the position of the gene responsible for prostate cancer was 15 further refined using the very high density set of markers described above. The results of this localization are shown in Figure 3.

As shown in Figure 3, the second phase of the analysis confirmed that the gene responsible for prostate cancer was near the biallelic marker designated 4-67, most probably within a ca. 150kb 20 region comprising the marker.

Haplotype analysis

The allelic frequencies of each of the alleles of biallelic markers 99-123, 4-26, 4-14, 4-77, 99-217, 4-67, 99-213, 99-221, and 99-135 (SEQ ID NOs: 21-38) were determined in the affected and unaffected populations. Table 1 lists the internal identification numbers of the markers used in the 25 haplotype analysis (SEQ ID NOs: 21-38), the alleles of each marker, the most frequent allele in both unaffected individuals and individuals suffering from prostate cancer, the least frequent allele in both unaffected individuals and individuals suffering from prostate cancer, and the frequencies of these alleles in each population.

Among all the theoretical potential different haplotypes based on 2 to 9 markers, 11 30 haplotypes showing a strong association with prostate cancer were selected. The results of these haplotype analyses are shown in Figure 4.

Figures 2, 3, and 4 aggregate linkage analysis results with sequencing results which permitted the physical order and/or the distance between markers to be estimated.

The significance of the values obtained in Figure 4 are underscored by the following results of computer simulations. For the computer simulations, the data from the affected individuals and the unaffected controls were pooled and randomly allocated to two groups which contained the same

number of individuals as the affected and unaffected groups used to compile the data summarized in Figure 4. A haplotype analysis was run on these artificial groups for the six markers included in haplotype 5 of Figure 4. This experiment was reiterated 100 times and the results are shown in Figure 5. Among 100 iterations, only 5% of the obtained haplotypes are present with a p-value below $1X10^{-4}$ as compared to the p-value of $9X10^{-7}$ for haplotype 5 of Figure 4. Furthermore, for haplotype 5 of Figure 4, only 6% of the obtained haplotypes have a significance level below $5X10^{-3}$, while none of them show a significance level below $5X10^{-5}$.

Thus, using the data of Figure 4 and evaluating the associations for single maker alleles or for haplotypes will permit estimation of the risk a corresponding carrier has to develop prostate cancer. Significance thresholds of relative risks will be adapted to the reference sample population used.

The diagnostic techniques may employ a variety of methodologies to determine whether a test subject has a biallelic marker pattern associated with an increased risk of developing prostate cancer or suffers from prostate cancer resulting from a mutant PG1 allele. These include any method enabling the analysis of individual chromosomes for haplotyping, such as family studies, single sperm DNA analysis or somatic hybrids.

In each of these methods, a nucleic acid sample is obtained from the test subject and the biallelic marker pattern for one or more of the biallelic markers listed in Figures 4, 6A and 6B is determined. The biallelic markers listed in Figure 6A are those which were used in the haplotype analysis of Figure 4. The first column of Figure 6A lists the BAC clones in which the biallelic markers lie. The second column of Figure 6A lists the internal identification number of the marker. The third column of Figure 6A lists the sequence identification number for a first allele of the biallelic markers. The fourth column of Figure 6A lists the sequence identification number for a second allele of the biallelic marker. For example, the first allele of the biallelic marker 99-123 has the sequence of SEQ ID NO: 30.

The fifth column of Figure 6A lists the sequences of upstream primers which is used to generate amplification products containing the polymorphic bases of the biallelic markers. The sixth column of Figure 6A lists the sequence identification numbers for the upstream primers.

The seventh column of Figure 6A lists the sequences of downstream primers which is used to generate amplification products containing the polymorphic bases of the biallelic markers. The eighth column of Figure 6A lists the sequence identification numbers for the downstream primers.

The ninth column of Figure 6A hists the position of the polymorphic base in the amplification products generated using the upstream and downstream primers. The tenth column lists the identities of the polymorphic bases found at the polymorphic positions in the biallelic markers. The eleventh and twelfth columns list the locations of microsequencing primers in the biallelic markers which can be used to determine the identities of the polymorphic bases.

10

In addition to the biallelic markers of SEQ ID NOs: 21-38, other biallelic markers (designated 99-1482, 4-73, 4-65) have been identified which are closely linked to one or more of the biallelic markers of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62, and the PG1 gene. These biallelic markers include the markers of SEQ ID NOs: 57-62, which are listed in Figure 6B. The columns in Figure 6B are identical to the corresponding columns in Figure 6A. SEQ ID NOs: 58, 59, 61, and 62 lie within the PG1 gene of SEQ ID NO:1 at the positions indicated in the accompanying Sequence Listing.

Genetic analysis of these additional biallelic markers is performed as follows. Nucleic acid samples are obtained from individuals suffering from prostate cancer and unaffected individuals. The frequencies at which each of the two alleles occur in the affected and unaffected populations is determined using the methodologies described above. Association values are calculated to determine the correlation between the presence of a particular allele or spectrum of alleles and prostate cancer. The markers of SEQ ID NOs: 21-38 may also be included in the analysis used to calculate the risk factors. The markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 is used in diagnostic techniques, such as those described below, to determine whether an individual is at risk for developing prostate cancer or suffers from prostate cancer as a result of a mutation in the PG1 gene.

Example 6 describes methods for determining the biallelic marker pattern.

Example 6

A nucleic acid sample is obtained from an individual to be tested for susceptibility to prostate cancer or PG1 mediated prostate cancer. The nucleic acid sample is an RNA sample or a DNA sample.

A PCR amplification is conducted using primer pairs which generate amplification products containing the polymorphic nucleotides of one or more biallelic markers associated with prostate cancer-related forms of PG1, such as the biallelic markers of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62, biallelic markers which are in linkage disequilibrium with the biallelic markers of SEQ ID NOs: 25 21-38, SEQ ID NOs: 57-62, biallelic markers in linkage disequilibrium with the PG1 gene, or combinations thereof. In some embodiments, the PCR amplification is conducted using primer pairs which generate amplification products containing the polymorphic nucleotides of several biallelic markers. For example, in one embodiment, amplification products containing the polymorphic bases of several biallelic markers selected from the group consisting of SEQ ID NOs: 21-38, SEQ ID NOs: 30 57-62, and biallelic markers which are in linkage disequilibrium with the biallelic markers of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62 or with the PG1 gene is generated. In another embodiment, amplification products containing the polymorphic bases of two or more biallelic markers selected from the group consisting of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62, and biallelic markers which are in linkage disequilibrium with the biallelic markers of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62 35 or with the PG1 gene is generated. In another embodiment, amplification products containing the polymorphic bases of five or more biallelic markers selected from the group consisting of SEQ ID

NOs: 21-38, SEQ ID NOs: 57-62, and biallelic markers which are in linkage disequilibrium with the biallelic markers of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62 or with the PG1 gene is generated. In another embodiment, amplification products containing the polymorphic bases of more than five of the biallelic markers selected from the group consisting of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62, and biallelic markers which are in linkage disequilibrium with the biallelic markers of SEQ ID NOs: 21-38, SEQ ID NOs: 57-62 or with the PG1 gene is generated.

For example, the primers used to generate the amplification products may comprise the primers listed in Figure 6A or 6B (SEQ ID NOs: 39-56 and SEQ ID NOs: 63-68). Figures 6A and Figure 6B provide exemplary primers which is used in the amplification reactions and the identities and locations of the polymorphic bases in the amplification products which are produced with the exemplary primers. The sequences of each of the alleles of the biallelic markers resulting from amplification using the primers in Figures 6A and 6B are listed in the accompanying Sequence Listing as SEQ ID NOs:21-38 and 57-62.

The PCR primers is oligonucleotides of 10, 15, 20 or more bases in length which enable the amplification of the polymorphic site in the markers. In some embodiments, the amplification product produced using these primers is at least 100 bases in length (i.e. 50 nucleotides on each side of the polymorphic base). In other embodiments, the amplification product produced using these primers is at least 500 bases in length (i.e. 250 nucleotides on each side of the polymorphic base). In still further embodiments, the amplification product produced using these primers is at least 1000 bases in length (i.e. 500 nucleotides on each side of the polymorphic base).

It will be appreciated that the primers listed in Figure 6A and 6B are merely exemplary and that any other set of primers which produce amplification products containing the polymorphic nucleotides of one or more of the biallelic markers of SEQ ID NOs. 21-38 and SEQ ID NOs: 57-62 or biallelic markers in linkage disequilibrium with the sequences of SEQ ID NOs. 21-38 and SEQ ID NOs: 57-62 or with the PG1 gene, or a combination thereof is used in the diagnostic methods.

Following the PCR amplification, the identities of the polymorphic bases of one or more of the biallelic markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62, or biallelic markers in linkage disequilibrium with the sequences of SEQ ID NOs. 21-38 and SEQ ID NOs: 57-62 or with the PG1 gene, or a combination thereof, are determined. The identities of the polymorphic bases is determined using the microsequencing procedures described in Example 5 above and the microsequencing primers listed as features in the sequences of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62. It will be appreciated that the microsequencing primers listed as features in the sequences of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 are merely exemplary and that any primer having a 3Nend near the polymorphic nucleotide, and preferably immediately adjacent to the polymorphic nucleotide, is used.

Alternatively, the microsequencing analysis is performed as described in Pastinen et al., Genome Research 7:606-614 (1997), which is described in more detail below.

15

Alternatively, the PCR product is completely sequenced to determine the identities of the polymorphic bases in the biallelic markers. In another method, the identities of the polymorphic bases in the biallelic markers is determined by hybridizing the amplification products to microarrays containing allele specific oligonucleotides specific for the polymorphic bases in the biallelic markers.

The use of microarrays comprising allele specific oligonucleotides is described in more detail below.

It will be appreciated that the identities of the polymorphic bases in the biallelic markers is determined using techniques other than those listed above, such as conventional dot blot analyses.

Nucleic acids used in the above diagnostic procedures may comprise at least 10 consecutive nucleotides in the biallelic markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 or the sequences complementary thereto. Alternatively, the nucleic acids used in the above diagnostic procedures may comprise at least 15 consecutive nucleotides in the biallelic markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 or the sequences complementary thereto. In some embodiments, the nucleic acids used in the above diagnostic procedures may comprise at least 20 consecutive nucleotides in the biallelic markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 or the sequences complementary thereto. In still other embodiments, the nucleic acids used in the above diagnostic procedures may comprise at least 30 consecutive nucleotides in the biallelic markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 or the sequences complementary thereto. In further embodiments, the nucleic acids used in the above diagnostic procedures may comprise more than 30 consecutive nucleotides in the biallelic markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 or the sequences complementary thereto. In still further embodiments, the nucleic acids used in the above diagnostic procedures may comprise the entire sequence of the biallelic markers of SEQ ID NOs: 21-38 and SEQ ID NOs: 57-62 or the sequences complementary thereto.

I.E. Identification and Sequencing, of the PG1 Gene, and Localization of the PG1 Protein

The above haplotype analysis indicated that 171kb of genomic DNA between biallelic markers 4-14 and 99-221 totally or partially contains a gene responsible for prostate cancer. Therefore, the protein coding sequences lying within this region were characterized to locate the gene associated with prostate cancer. This analysis, described in further detail below, revealed a single protein coding sequence in the 171 kb, which was designated as the PG1 gene.

Template DNA for sequencing the PG1 gene was obtained as follows. BACs 189EO8 and 30 463FO1 were subcloned as previously described Plasmid inserts were first amplified by PCR on PE 9600 thermocyclers (Perkin-Elmer), using appropriate primers, AmpliTaqGold (Perkin-Elmer), dNTPs (Boehringer), buffer and cycling conditions as recommended by the Perkin-Elmer Corporation.

PCR products were then sequenced using automatic ABI Prism 377 sequencers (Perkin Elmer, Applied Biosystems Division, Foster City, CA). Sequencing reactions were performed using PE 9600 35 Chermocyclers (Perkin Elmer) with standard dye-primer chemistry and ThermoSequenase (Amersham Life Science). The primers were labeled with the JOE, FAM, ROX and TAMRA dyes. The dNTPs and

ddNTPs used in the sequencing reactions were purchased from Boehringer. Sequencing buffer, reagent concentrations and cycling conditions were as recommended by Amersham.

41

Following the sequencing reaction, the samples were precipitated with EtOH, resuspended in formamide loading buffer, and loaded on a standard 4% acrylamide gel. Electrophoresis was performed for 2.5 hours at 3000V on an ABI 377 sequencer, and the sequence data were collected and analyzed using the ABI Prism DNA Sequencing Analysis Software, version 2.1.2.

The sequence data obtained as described above were transferred to a proprietary database, where quality control and validation steps were performed. A proprietary base-caller ("Trace"), working using a Unix system automatically flagged suspect peaks, taking into account the shape of the peaks, the interpeak resolution, and the noise level. The proprietary base-caller also performed an automatic trimming. Any stretch of 25 or fewer bases having more than 4 suspect peaks was considered unreliable and was discarded. Sequences corresponding to cloning vector oligonucleotides were automatically removed from the sequence. However, the resulting sequence may contain 1 to 5 bases belonging to the vector sequences at their 5• end. If needed, these can easily be removed on a case by case basis.

The genomic sequence of the PG1 gene is provided in the accompanying Sequence Listing and is designated as SEQ ID NO: 1.

Potential exons in BAC-derived human genomic sequences were located by homology searches on protein, nucleic acid and EST (Expressed Sequence Tags) public databases. Main public databases were locally reconstructed. The protein database, NRPU (Non-redundant Protein Unique) is formed by a non-redundant fusion of the Genpept (Benson D.A. et al., Nucleic Acids Res. 24: 1-5 (1996), Swissprot (Bairoch, A. and Apweiler, R, Nucleic Acids Res. 24: 21-25 (1996) and PIR/NBRF (George, D.G. et al., Nucleic Acids Res. 24:17-20 (1996) databases. Redundant data were eliminated by using the NRDB software (Benson et al., supra) and internal repeats were masked with the XNU software (Benson et al., supra). Homologies found using the NRPU database allowed the identification of sequences corresponding to potential coding exons related to known proteins.

The EST local database is composed by the gbest section (1-9) of GenBank (Benson et al., supra), and thus contains all publicly available transcript fragments. Homologies found with this database allowed the localization of potentially transcribed regions.

The local nucleic acid database contained all sections of GenBank and EMBL (Rodriguez-Tome, P. et al., Nucleic Acids Res. 24: 6-12 (1996) except the EST sections. Redundant data were eliminated as previously described.

Similarity searches in protein or nucleic acid databases were performed using the BLAS software (Altschul, S.F. et al., J. Mol. Biol. 215: 403-410 (1990). Alignments were refined using the Fasta software, and multiple alignments used Clustal W. Homology thresholds were adjusted for each analysis based on the length and the complexity of the tested region, as well as on the size of the reference database.

Potential exon sequences identified as above were used as probes to screen cDNA libraries. Extremities of positive clones were sequenced and the sequence stretches were positioned on the genomic sequence of SEQ ID NO:1. Primers were then designed using the results from these alignments in order to enable the PG1 cloning procedure described below.

Cloning PG1 cDNA

PG1 cDNA was obtained as follows. 4:1 of ethanol suspension containing 1mg of human prostate total RNA (Clontech laboratories, Inc., Palo Alto, USA; catalogue N. 64038-1, lot 7040869) was centrifuged, and the resulting pellet was air dried for 30 minutes at room temperature.

First strand cDNA synthesis was performed using the AdvantageTM RT-for-PCR kit (Clontech laboratories, Inc., Palo Alto, USA; catalogue N. K1402-1). 1:l of 20mM solution of primer PGRT32: TTTTTTTTTTTTTTTTGAAAT (SEQ ID NO:10) was added to 12.5 :1 of RNA solution in water, heated at 74°C for two and a half minutes and rapidly quenched in an ice bath. 10:1 of 5xRT buffer (50mM Tris-HCl, pH 8.3, 75mM KCl, 3 mM MgCl2), 2.5 :l of dNTP mix (10mM each), 1.25:1 of human recombinant placental RNA inhibitor were mixed with 1 ml of MMLV reverse 15 transcriptase (200 units). 6.5:1 of this solution were added to RNA-primer mix and incubated at 42°C for one hour. 80:1 of water were added and the solution was incubated at 94°C for 5 minutes.

5:1 of the resulting solution were used in a Long Range PCR reaction with hot start, in 50:1 pmol/µl of each of GC1.5p.1: using 2 units of rtTHXL, 20 final volume. ID NO:6) or GC1.5p2 CTGTCCCTGGTGCTCCACACGTACTC (SEQ ID NO: 7) and GC1.3p: 20 TGGTGCTCCACACGTACTCCATGCGC (SEQ CTTGCCTGCTGGAGACACAGAATTTCGATAGCAC (SEQ ID NO:9) primers with 35 cycles of elongation for 6 minutes at 67°C in thermocycler.

The sequence of the PG1 cDNA obtained as described above (SEQ ID NO 3) is provided in the accompanying Sequence Listing. Results of Northern blot analysis of prostate mRNAs support the 25 existence of a major PG1 cDNA having a 5-6kb length.

Characterization of the PG1 Gene

The intron/exon structure of the gene was deduced by aligning the mRNA sequence from the cDNA of SEQ ID NO:3 and the genomic DNA sequence of SEQ ID NO: 1.

The positions of the introns and exons in the PG1 genomic DNA are provided in Figures 7 and 8. Figure 7 lists positions of the start and end nucleotides defining each of the at least 8 exons (labeled Exons A-H) in the sequence of SEQ ID NO: 1, the locations and phases of the 5' and 3' splice sites in the sequence of SEQ ID NO: 1, the position of the stop codon in the sequence of SEQ ID NO: 1, and the position of the polyadenylation site in the sequence of SEQ ID NO: 1. Figure 8 shows the positions of the exons within the PG1 genomic DNA and the PG1 mRNA, the location of a 35 tyrosine phosphatase retro-pseudogene in the PG1 genomic DNA, the positions of the coding region in the mRNA, and the locations of the polyadenylation signal and polyA stretch in the mRNA.

5

As indicated in Figures 7 and 8, the PG1 gene comprises at least 8 exons, and spans more than 52kb. The first intron contains a tyrosine phosphatase retropseudogene. A G/C rich putative promoter region lies between nucleotide 1629 and 1870 of SEQ ID NO: 1. A CCAAT box is present at nucleotide 1661 of SEQ ID NO: 1. The promoter region was identified as described in Prestridge, D.S., Predicting Pol II Promoter Sequences Using Transcription Factor Binding Sites, J. Mol. Biol. 249:923-932 (1995).

It is possible that the methionine listed as being the initiating methionine in the PG1 protein sequence of SEQ ID NO: 4 (based on the cDNA sequence of SEQ ID NO: 3) may actually be downstream but in phase with another methionine which acts as the initiating methionine. The genomic DNA sequence of SEQ ID NO: 1 contains a methionine upstream from the methionine at position number 1 of the protein sequence of SEQ ID NO: 4. If the upstream methionine is in fact the authentic initiation site, the sequence of the PG1 protein would be that of SEQ ID NO: 5. This possibility is investigated by determining the exact position of the 5N end of the PG1 mRNA as follows.

One way to determine the exact position of the 5N end of the PG1 mRNA is to perform a 5NRACE reaction using the Marathon-Ready human prostate cDNA kit from Clontech (Catalog. No. PT1156-1). For example, the RACE reaction may employ the PG1 primers PG15RACE196 CAATATCTGGACCCCGGTGTAATTCTC (SEQ ID NO: 8) as the first primer. The second primer sequence the PG15RACE130n having is reaction **RACE** in the GGTCGTCCAGCGCTTGGTAGAAG (SEQ ID NO: 2). The sequence analysis of the resulting PCR product, or the product obtained with other PG1 specific primers, will give the exact sequence of the initiation point of the PG1 transcript.

Alternatively, the 5Nsequence of the PG1 transcript can be determined by conducting a PCR amplification with a series of primers extending from the 5Nend of the presently identified coding region. In any event, the present invention contemplates use of PG1 nucleic acids and/or polypeptides coding for or corresponding to either SEQ ID NO:4 or SEQ ID NO:5 or fragments thereof.

It is also possible that alternative splicing of the PG1 gene may result in additional translation products not described above. It is also possible that there are sequences upstream or downstream of the genomic sequence of SEQ ID NO: 1 which contribute to the translation products of the gene. Finally, alternative promoters may result in PG1derived transcripts other than those described herein.

The promoter activity of the region between nucleotides 1629 and 1870 can be verified as described below. Alternatively, should this region lack promoter activity, the promoter responsible for driving expression of the PG1 gene is identified as described below.

Genomic sequences lying upstream of the PG1 gene are cloned into a suitable promoter reporter vector, such as the pSEAP-Basic, pSEAP-Enhancer, pβgal-Basic, pβgal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech. Briefly, each of these promoter reporter vectors

15

20

include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase, β galactosidase, or green fluorescent protein. The sequences upstream of the PG1 coding region are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert. If necessary, the upstream sequences can be cloned into vectors which contain an enhancer for augmenting transcription levels from weak promoter sequences. A significant level of expression above that observed with the vector lacking an insert indicates that a promoter sequence is present in the inserted upstream sequence.

Promoter sequences within the upstream genomic DNA is further defined by constructing nested deletions in the upstream DNA using conventional techniques such as Exonuclease III digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has reduced or obliterated promoter activity. In this way, the boundaries of the promoters is defined. If desired, potential individual regulatory sites within the promoter is identified using site directed mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter individually or in combination. The effects of these mutations on transcription levels is determined by inserting the mutations into the cloning sites in the promoter reporter vectors.

Sequences within the PG1 promoter region which are likely to bind transcription factors is identified by homology to known transcription factor binding sites or through conventional mutagenesis or deletion analyses of reporter plasmids containing the promoter sequence. For example, deletions is made in a reporter plasmid containing the promoter sequence of interest operably linked to an assayable reporter gene. The reporter plasmids carrying various deletions within the promoter region are transfected into an appropriate host cell and the effects of the deletions on expression levels is assessed.

Transcription factor binding sites within the regions in which deletions reduce expression levels is further localized using site directed mutagenesis, linker scanning analysis, or other techniques familiar to those skilled in the art.

The promoters and other regulatory sequences located upstream of the PG1 gene is used to design expression vectors capable of directing the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative manner. For example, since the PG1 promoter is presumably active in the prostate, it can be used to construct expression vectors for directing gene expression in the prostate.

Preferably, in such expression vectors, the PG1 promoter is placed near multiple restriction sites to facilitate the cloning of an insert encoding a protein for which expression is desired downstream of the promoter, such that the promoter is able to drive expression of the inserted gene. The promoter is inserted in conventional nucleic acid backbones designed for extrachromosomal replication, integration

into the host chromosomes or transient expression. Suitable backbones for the present expression vectors include retroviral backbones, backbones from eukaryotic episomes such as SV40 or Bovine Papilloma Virus, backbones from bacterial episomes, or artificial chromosomes.

Preferably, the expression vectors also include a polyA signal downstream of the multiple restriction sites for directing the polyadenylation of mRNA transcribed from the gene inserted into the expression vector.

Nucleic acids encoding proteins which interact with sequences in the PG1 promoter is identified using one-hybrid systems such as those described in the manual accompanying the Matchmaker One-Hybrid System kit available from Clontech (Catalog No. K1603-1). Briefly, the Matchmaker One-hybrid system is used as follows. The target sequence for which it is desired to identify binding proteins is cloned upstream of a selectable reporter gene and integrated into the yeast genome. Preferably, multiple copies of the target sequences are inserted into the reporter plasmid in tandem.

A library comprised of fusions between cDNAs to be evaluated for the ability to bind to the promoter and the activation domain of a yeast transcription factor, such as GALA, is transformed into the yeast strain containing the integrated reporter sequence. The yeast are plated on selective media to select cells expressing the selectable marker linked to the promoter sequence. The colonies which grow on the selective media contain genes encoding proteins which bind the target sequence. The inserts in the genes encoding the fusion proteins are further characterized by sequencing. In addition, the inserts is inserted into expression vectors or in vitro transcription vectors. Binding of the polypeptides encoded by the inserts to the promoter DNA is confirmed by techniques familiar to those skilled in the art, such as gel shift analysis or DNAse protection analysis.

Analysis of PG1 Protein Sequence

The PG1 cDNA of SEQ ID NO: 3 encodes a 353 amino-acid protein (SEQ ID NO:4). As indicated in the accompanying Sequence Listing, a Prosite analysis indicated that the PG1 protein has a leucine zipper motif, a potential glycosylation site, 3 potential casein kinase II phosphorylation sites, a potential cAMP dependent protein kinase phosphorylation site, 2 potential tyrosine kinase phosphorylation sites, 4 potential protein kinase C phosphorylation sites, 5 potential N-myristoylation sites, 1 potential tyrosine sulfation site, and one potential amidation site.

A search for membrane associated domains was conducted according to the methods described in Argos, P. et al., Structural Prediction of Membrane-bound Proteins, Elur. J. Biochem. 128:565-575 (1982); Klein et al., Biochimica & Biophysica Acta 815:468-476 (1985); and Eisenberg et al., J. Mol. Biol. 179:125-142 (1984). The search revealed 5 potential transmembrane domains predicted to be integral membrane domains. These results suggest that the PG1 protein is likely to be membrane-associated and is an integral membrane protein.

A homology search was conducted to identify proteins homologous to the PG1 protein. Several proteins were identified which share homology with the PG1 protein. Figure 9 lists the

accession numbers of several proteins which share homology with the PG1 protein in three regions designated box1, box2 and box3.

It will be appreciated that each of the motifs described above is also present in the protein of SEQ ID NO: 5, which would be produced if by translation initiation translated from the potential 5 upstream methionine in the nucleic acid of SEQ ID NO: 1.

As indicated in Figure 9, a distinctive pattern of homology to box 1, box 2 (SEQ ID NOs: 11-14) and box 3 (SEQ ID NOs: 15-20) is found amongst acyl glyerol transferases. For example, the plsC protein from E. coli (Accession Number P26647) shares homology with the box1 and box2 sequences, but not the box 3 sequence, of the PG1 protein. The product of this gene transfers acyl 10 from acyl-coenzymeA to the sn2 position of 1-Acyl-sn-glycerol-3-phosphate (lysophosphatidic acid, LPA)(Coleman J., Mol Gen Genet. 1992 Mar 1; 232(2): 295-303).

Box1 and box2 homologies, but not box 3 homologies, are also found in the SLCI gene product from baker's yeast (Accession Number P33333) and the mouse gene AB005623. Each of these genes are able to complement in vivo mutations in the bacterial plsC gene. (Nagiec MM, Wells 15 GB, Lester RL, Dickson RC, J. Biol. Chem., 1993 Oct 15; 268(29): 22156-22163, A suppressor gene that enables Saccharomyces cerevisiae to grow without making sphingolipids encodes a protein that resembles an Escherichia coli fatty acyltransferase; and Kume K, Shimizu T, Biochem. Biophys. Res. Commun. 1997, Aug. 28; 237(3): 663-666, cDNA cloning and expression of murine 1-acyl-snglycerol-3-phosphate acyltransferase).

Recently two different human homologues of the mouse AB005623 gene, Accession Numbers U89336 and U56417 were cloned and found to be localized to human chromosomes 6 and 9 (Eberhardt. C., Gray, P.W. and Tjoelker, L.W., J. Biol. Chem. 1997; 272, 20299-20305, Human lysophosphatidic acid acyltransferase cDNA cloning, expression, and localization to chromosome 9q34.3; and West, J., Tompkins, C.K., Balantac, N., Nudelman, E., Meengs, B., White, T., Bursten, 25 S., Coleman, J., Kumar, A., Singer, J.W. and Leung, D.W, DNA Cell Biol. 6, 691-701 (1997), Cloning and expression of two human lysophosphatidic acid acyltransferase cDNAs that enhance cytokine induced signaling responses in cells).

The enzymatic acylation of LPA results in 1,2-diacyl-sn-glycerol 3-phosphate, an intermediate to the biosynthesis of both glycerophospholipids and triacylglycerol. Several important signaling 30 messengers participating in the transduction of mitogenic signals, induction of apoptosis, transmission of nerve impulses and other cellular responses mediated by membrane bound receptors belong to this metabolic pathway.

LPA itself is a potent regulator of mammalian cell proliferation. In fact, LPA is one of the major mitogens found in blood serum. (For a review: Durieux ME, Lynch KR, Trends Pharmacol. 35 Sci. 1993 Jun; 14(6):249-254, Signaling properties of lysophosphatidic acid. LPA can act as a survival factor to inhibit apoptosis of primary cells; and Levine JS, Koh JS, Triaca V, Lieberthal W,

WO 99/32644 PCT/IB98/02133

Am. J. Physiol. 1997 Oct; 273(4Pt2): F575-F585, Lysophosphatidic acid: a novel growth and survival factor for renal proximal tubular cells). This function of LPA is mediated by the lipid kinase phosphatidylinositol 3-kinase.

Phosphatidylinositol and its derivatives present another class of messengers emerging from the 1-acyl-sn-glycerol-3-phosphate acyltransferase pathway. (Toker A, Cantley LC, Nature 1997 Jun 12; 387(6634): 673-676, Signaling through the lipid products of phosphoinositide-3-OH kinase; Martin TF, Curr. Opin. Neurobiol. 1997 Jun; 7(3):331-338, Phosphoinositides as spatial regulators of membrane traffic; and Hsuan JJ, et al., Int. J. Biochem. Cell Biol. 1997 Mar 1st; 29(3): 415-435, Growth factor-dependent phosphoinositide signaling).

Cell growth, differentiation and apoptosis can be affected and modified by enzymes involved in this metabolic pathway. Consequently, alteration of this pathway could facilitate cancer cell progression. Modulation of the activity of enzymes in this pathway using agents such as enzymatic inhibitors could be a way to restore a normal phenotype to cancerous cells.

Ashagbley A, Samadder P, Bittman R, Erukulla RK, Byun HS, Arthur G have recently shown that ether-linked analogue of lysophosphatidic acid: 4-O-hexadecyl-3(S)-O-methoxybutanephosphonate can effectively inhibit the proliferation of several human cancerous cell lines, including DU145 line of prostate cancer origin. (Anticancer Res 1996 Jul; 16(4A): 1813-1818, Synthesis of ether-linked analogues of lysophosphatidate and their effect on the proliferation of human epithelial cancer cells in vitro).

Structural differences between the PG1 family of cellular proteins and the functionally confirmed 1-acyl-sn-glycerol-3-phosphate acyltransferase family, evidenced by the existence of a different pattern of homology to box3, could point to unique substrate specificity in the phospholipid metabolic pathway, to specific interaction with other cellular components or to both.

Further analysis of the function of the PG1 gene can be conducted, for example, by constructing knockout mutations in the yeast homologues of the PG1 gene in order to elucidate the potential function of this protein family, and to test potential substrate analogs in order to revert the malignant phenotype of human prostate cancer cells as described in Section VIII, below.

Example 7

Analysis of the Intracellular Localisation of the PG1 Isoforms

To study the intracellular localisation of PG1 protein, different isoforms of PG1 were cloned in the expression vector pEGFP-N1(Clontech), transfected and expressed in normal (PNT2A) or adenocarcinoma (PC3) prostatic cell line.

First, to generate cDNA inserts, 5' and 3' primers were synthesised allowing to amplify different regions of the PG1 open reading frame. Respectively, these primers were designed with an internal EcoRI or BamHI site which allowed the insertion of the amplified product into the EcoRI and BamHI sites of the expression vector. The restriction sites were introduced into the primer so that

10

15

after cloning into pEGFP-N1, the PG1 open reading frame would be fused in frame, to the EGFP open reading frame. The translated protein would be a fusion between PG1 and EGFP. EGFP being a variant form of the GFP protein (Green Fluorescent Protein), it is possible to detect the intracellular localisation of the different PG1 isoforms by examining the fluorescence emitted by the EGFP fused protein.

The different forms that were analysed correspond either to different messengers identified by RT-PCR performed on total normal human prostatic RNA or to a truncated form resulting from a non sense mutation identified in a tumoural prostatic cell line LnCaP. The different PG1 constructions were transfected using the lipofectine technique and EGFP expression was examined 20 hours post transfection.

Name and description of the different forms transfected are listed below:

- A) PG1 includes all the coding exons from exon 1 to 8.
- B) PG1/1-4 corresponds to an alternative messenger which is due to an alternative splicing, joining exon 1 to exon 4, and resulting in the absence of exons 2 and 3.
- 15 C) PG1/1-5 corresponds to an alternative messenger which is due to an alternative splicing, joining exon 1 to exon 5, and resulting in the absence of exons 2, 3 and 4.
 - D) PG1/1-7 includes exons 1 to 6, and corresponds to the mutated form identified in genomic DNA of the prostatic tumoural cell line LNCaP.

Cloning of the PG1 cDNA inserts in the EGFP-N1 expression vector

cDNAs from human prostate were obtained by RT-PCR using the Advantage RT-for-PCR Kit 20 (CLONTECH ref K1402-2). First, 1µl of oligodT-containing PG1 specific primer PGRT32 TTTTTTTTTTTTTTTGAAAT (20pmoles) and 11.5 μl of DEPC treated H₂O were added to lµl of total mRNA (lµg) extracted from human prostate (CLONTECH ref 64038-1). The mRNA was heat denaturated for 2.5 min at 74°C and then quickly chilled on ice. A mix containing 4µl of 5X buffer, 1µl of dNTPs (10mM each), 0.5µl of recombinant RNase inhibitor (20U) and 1µl of MoMuLV Reverse Transcriptase (200U) was added to the denaturated mRNA. Reverse transcription was performed for 60 min. at 42°C. Enzymes were heat denaturated for 5 min. at 94°C. Then, 80µl of DEPC treated H₂O were added to the reaction mix and the cDNA mix was stored at -20°C. Primers CCTGAATTCCGCCGAGCTGAGAAGATGC 3'), and PG13Bam2 (5' PG15Eco3 30 CCTGGATCCGCTTTAATAGTAACCCACAGGCAG 3') were used for PCR amplification of the different PG1 cDNAs. A 50µl PCR reaction mix containing 5µl of the previously prepared prostate cDNA mix, 15µl of 3.3X PCR buffer, 4µl of dNTPs (2.5mM each), 20pmoles of primer PG15Eco3, 20pmoles of primer PG13Bam2, 1μl of RtthXL enzyme, 2.2μl Mg(OAc)₂ (Hot Start) was set up and amplification was performed for 35 cycles of 30 sec at 94°C, 10 min. at 72°C, 4 min. at 67°C after an 35 initial denaturation step of 10 min. at 94°C. Size and integrity of the PCR product was assessed by migration on a 1% agarose gel. 2µg of the amplification product were digested with 2.4 units of

20

30

EcoRI (PROMEGA ref R601A) and 2.0 units of BamHI (PROMEGA ref R602A) in 50µl of 1X Multicore buffer for 2 hours at 37°C. Enzymes were then heat inactivated for 20 min, at 68°C, DNA was phenol/chloroform extracted and ethanol-precipitated and its concentration was estimated by migration on a 1% agarose gel.

To prepare the vector, 2µg of pEGFP-N1 vector (CLONTECH ref 6085-1) were digested with 2.4 units of EcoRI (PROMEGA ref R601A) and 2.0 units of BamHI (PROMEGA ref R602A) in 50µl of 1X multicore buffer for 2 hours at 37°C. Enzymes were then heat inactivated for 20 min, at 68°C, DNA was phenol/chloroform extracted and ethanol-precipitated and its concentration and integrity were estimated by migration on a 1% agarose gel. 20ng of the BamHI and EcoRI digested pEGFP-N1 10 vector were added to 50ng of BamHI-EcoRI digested PG1 cDNAs. Ligation was performed over night at 13°C using 0.5units of T4 DNA ligase (BOEHRINGER ref 84333623) in a final volume of 20µl containing 1X ligase buffer. The ligation reaction mix was desalted by dialysis against water (MILLIPORE ref VSWP01300) for 30min. at room temperature. One fifth of the desalted ligation reaction was electroporated in 25µl of competent cells ElectroMAX DH10B (GIBCO BRL ref 15 18290-015) using a resistance of 126 Ohms, capacitance of 50μF, and voltage of 2.5KV. Bacteria were then incubated in 500µl of SOB medium for 30min at 37°C. One fifth was plated on LB AGAR containing 40µg/µl KANAMYCINE (SIGMA ref K4000) and incubated over night at 37°C.

Plasmid DNA was prepared from an overnight liquid culture of individual colonies and sequenced. Among the different forms identified 3 were used:

- A) PG1 which includes all the coding exons from exon 1 to 8.
- B) PG1/1-4 which corresponds to an alternative messenger which is due to an alternative splicing, joining exon 1 to exon 4, and resulting in the absence of exons 2 and 3.
- C) PG1/1-5 which corresponds to an alternative messenger which is due to an alternative splicing, joining exon 1 to exon 5, and resulting in the absence of exons 2, 3 and 4.
- D) Vector PG1/1-7: A cDNA insert encoding for a truncated protein was synthesized by PCR 25 (5' PG1mut29Bam PG15Eco3 and primers amplification, using CCTGGATCCCCTCCATCGTCTTTCCCTT 3') and vector PG1 as a template. The resulting PCR product was cloned following the same protocol as described above.

Transfection of the PG1 expression vectors in human prostate cell lines.

The DNA/lipofectin solution was prepared as followed: 1.5µl of lipofectin (GIBCO BRL ref 18292-011) was diluted in 100µl of OPTI-MEM medium (GIBCO BRL ref 31985-018), and incubated for 30min. at room temperature before being mixed to 0.5µg of vector diluted in 100µl of OPTI-MEM medium and incubated for 15 min. at room temperature. Cells were inoculated in RPMI1640 medium (Gibco BRL ref 61870-010) containing 5% fetal calf serum (Dutscher ref P30-35 3302) on slides (NUNC Lab-Tek ref 177402A) and grown at 37°C in 5%CO2. Cells reaching 40-60% confluency were rinsed with 300µl OPTI-MEM medium and incubated with the DNA/lipofectin solution for 6 hours at 37°C. The medium containing DNA was replaced by medium supplemented in fetal calf serum and cells were incubated for at least 36 hours at 37°C. Slides were rinsed in PBS and cells were fixed in ethanol, treated with Propidium iodide, and examined with a fluorescence microscope using a double-pass filter set for FITC/PI.

After transfection of PG1 and PG1/1-4 in both the normal and tumoural prostatic cell line, green fluorescence was detected into and around the nucleus (Figures 10 and 11). This result shows that the PG1 protein is localised in the nucleus and/or the nuclear membrane. Furthermore, it suggests that exons 2 and 3 are dispensable for translocation of PG1 to the nucleus. In addition, no difference in the intracellular localisation of these two forms was detected between the tumoral and the normal prostatic cell line.

On the contrary, transfection experiments using PG1/1-5 show that this form is cytoplasmic in the normal prostatic cell line PNT2A. It suggests that exon 4 might be important for the regulation of the translocation to the nucleus. Interestingly, similar transfection experiments in the tumoral cell line PC3 show that PG1/1-5 remains nuclear and or perinuclear (Figure 12). This result shows that there is 15 an abnormality in the regulation of the intracellular localization of the PG1 isoforms in this tumoral cell line. Furthermore, it indicates that the normal function of PG1 can be altered indirectly in prostatic tumors by an abnormality in the regulation of its intracellular location.

Finally, a non-sense mutation has been identified in the prostatic tumoural cell line LNCaP, in exon 6 of PG1 (SEQ ID NO: 69). This mutation is responsible for the production of a truncated 20 protein (SEQ ID NO: 70). To determine the intracellular location of this truncated protein, PG1/1-7 and PG1 were transfected in the normal prostatic cell line PNT2A. Comparison of the fluorescence detected in both sets of experiments clearly showed that the truncated form was localised in the cytoplasm as the non-truncated protein was located in and/or around the nucleus (Figure 13). This result indicates that this mutated PG1 is translated in a truncated protein which is unable to reach the nucleus. It also suggests that exons 7 and 8 may play an important role in the regulation of the intracellular localisation of PG1. Furthermore, it supports the previous hypothesis that an altered regulation of PG1 intracellular localisation might be involved in prostate tumorigenesis.

	pEGFP N1	PG1	PG11-4	PG11-5	PG11-7
Transfection PNT2 06/17/98	NA	nuclear •	nuclear	ND	ND
Transfection PNT2 06/30/98	cytoplasmic	nuclear	nuclear	ND	ND

5

X2-3-4Spliced out mut aa229

Transfection PNT2	cytoplasmic	NA	NA	cytoplasmic	ND
Transfection PC3 07/16/98	NA	nuclear	nuclear	nuclear	ND
Transfection PC3 07/16/98 bis	cytoplasmic	nuclear	NA	NA	ND
Transfection PC3 08/27/98	cytoplasmic	nuclear	nuclear	nuclear	NA
Transfection PNT2 08/28/98	cytoplasmic	nuclear	NA	cytoplasmic	cytoplasmic

X2-3 Spliced out

NA: Not assessable

ND: Not done

5 Nuclear: localized in and around the nucleus (nuclear and perinuclear localization).

All exons

Alternative Splice Species

Alternative splicing is a common natural tool for the inhibition of function of full length gene products. Alternative splicing is known to result in enzyme isoforms, possesing different kinetic characteristics (pyruvate kinase: M1 and M2 Yamada K, Noguchi T, Biochem J. 1999 Jan1;337(Pt 1):1-11. Estrogen receptor (ER) gene is known to possess variant splicing yelding the deletions of exon 3, 5, or 7. The truncated ER protein induced from variant mRNA could mainly be exhibited as a repressor through dominant negative effects on normal ER protein (Iwase H, Omoto Y, Iwata H, Hara Y, Ando Y, Kobayashi S, Oncology 1998 Dec;55 Suppl S1:11-16)'Yu et al (Yu JJ, Mu C, Dabholkar M, Guo Y, Bostick-Bruton F, Reed E,Int J Mol Med 1998 Mar;1(3):617-620) demonstrated that there is an association between alternative splicing of ERCC1, and reduction in cellular capability to repair cisplatin-DNA adduct. Munoz-Sanjuan et al (Munoz-Sanjuan I, Simandl BK, Fallon JF, Nathans J, Development 1998 Dec 14;126(Pt 2):409-421) demonstrated existence of two differentially spliced isoforms of fibroblast growth factor(FGF) type two genes that are present in non-overlapping spatial

distributions in the neural tube and adjacent structures in developing chiken embryo. One of these forms is secreted and activates the expression of HoxD13, HoxD11, Fgf-4 and BMP-2 ectopically, consistent with cFHF-2 playing a role in anterior-posterior patterning of the limb.

The CD44 is a cell adhesion molecule that is present as numerous isoforms created by mRNA alternative splicing. Expression of variant isoforms of CD44 is associated with tumor growth and metastasis. (Shibuya Y, Okabayashi T, Oda K, Tanaka N, Jpn J Clin Oncol 1998 Oct;28(10):609-14) they showed that ratio of two particular isoforms is a useful indicator of prognosis in gastric and colorectal carcinoma. Zhang YF et al (Zhang YF, Jeffery S, Burchill SA, Berry PA, Kaski JC, Carter ND, Br J Cancer 1998 Nov;78(9):1141-6° showed that human endothelin receptor A is the subject to alternative spicing giving at least two isoforms. The truncated receptor was expressed in all tissues and cells examined, but the level of expression varied. In melanoma cell lines and melanoma tissues, the truncated receptor gene was the major species, whereas the wild-type ETA was predominant in other tissues. Zhang et al. conclude that the function and biological significance of this truncated ETA receptor is not clear, but it may have regulatory roles for cell responses to ETs.

Example 8

Identification of PG1 Alternative Splice Species

The PG1 cDNA was first cloned by screening of a human prostate cDNA library. Sequence analysis of about 400 cDNA clones showed that at least 14 isoforms were present in this cDNA library. Comparison of their sequences to the genomic sequence showed that these isoforms resulted from a complex set of different alternative splicing events between numerous exons (Figure 14).

To rule out the possibility of a cloning artefact generated during the cDNA library construction and to systematically identify all existing alternative splice junctions, RT-PCR experiments were performed on RNA of normal prostate as well as normal prostatic cell lines PNT1A, PNT1B and PNT2 using all the possible combinations of primers specific to the different exon borders SEQ ID NOs: 137-178. The presence of multiple PCR bands in each reaction was assessed by migration in an agarose gel. Each band was analysed by sequencing, and the presence or absence of specific splicing events, as seen in the sequence by a specific splice junction, was scored as plus or minus in Figure 15.

Furthermore, to identify aberrant splicing event in prostate tumors, similar experiments were performed on RNA extracted from tumoral prostatic cell lines LnCaP (obtained from two different sources and named FCG and JMB), CaHPV, Du145 and PC3 as well as on RNA obtained from prostate tumors (ECP5 to ECP24).

As shown in the first five columns, all isoforms identified in the cDNA library were detected in RNA of normal prostate, normal prostatic cell lines or prostate tumors. In addition to the different splice junctions detected in the cDNA library, 19 other splice junctions were detected in normal prostate or in normal prostatic cell lines. Two types of exon junctions (exons 3-7, exons 3b-8) were

30

never detected in either normal prostate, normal prostatic cell lines, prostate tumors or prostatic tumoral cell lines. Comparison between normal and tumoral samples showed the presence of 2 additional exon junctions (exons 3-8, exons 5-8) in the tumoral samples that were not detected previously in the normal samples. This result demonstrate that during tumorigenesis, the complex regulation of the PG1 splicing has been altered, resulting in an abnormal ratio of the different isoforms. It is of a specific interest since it has been shown in patients with a genetic predisposition to Wilms tumor, that an imbalance between different RNA isoforms might be involved in tumorigenesis (Bickmore et al., Science 1992, 257:325-7; Little et al, Hum Mol Genet 1995, 4:351-8).

Interestingly, comparison between normal and tumoral samples, also showed that some exon junctions are present in all normal samples, but are absent in numerous tumoral samples. It further indicates that the normal function of PG1 can be altered by an abnormality in the regulation of PG1 splicing and further support the previous hypothesis.

Furthermore, comparison between the different types of normal samples (Col.2 versus Col. 3, 4 and 5) also showed differences in the presence or absence of some exon junctions. It indicates that the transformation process necessary to the generation of these normal prostatic cell lines might result in similar alteration which further support the previous hypothesis.

Example 9

Determining the Tumor Suppressor Activity of the PG1 Gene Product, Mutants and Other PG1 **Polypeptides**

PG1 variants which results from either alternate splicing of the PG1 mRNA or from mutation of PG1 that introduce a stop codon (nucleotide of SEQ ID NO: 69 and protein of SEQ ID NO: 70) can no longer perform its role of tumor suppressor. It is possible and even likely that PG1 tumor suppressor role extends beyond prostate cancer to other form of malignancies. PG1 therefore represent a prime candidate for gene therapy of cancer by creating a targeting vector which knocks out the mutant and/or introduces a wild-type PG1 gene (e.g. SEQ ID NO 3 or 179) or a fragment thereof.

To validate this model, PG1 and its alternatively spliced or mutated variants are stably transfected in tumor cell line using methods described in Section VIII. The efficiency of transfection is determined by northern and western blotting; the latter is performed using antibodies prepared against PG1 synthetic peptides designed to distinguish the product of the most abundant PG1 mRNA from the alternatively spliced variants, the truncated variant, or other functional mutants. The production of synthetic peptides and of polyclonal antibodies is performed using the methods described herein in Sections III and VII.. After demonstrating that PG1 and its variant are efficiently expressed in various tumor cell line preferably derived from human prostate cancer, hepatocarcinoma, lung and colon carcinoma; we the effect of this gene on the rate of cell division, DNA synthesis, ability to grow in soft agar and ability to induce tumor progression and metastasis when injected in immunologically deficient nude mice are determined.

35

20

Alternatively the PG1 gene and its variant are inserted in adenoviruses that are used to obtain a high level of expression of these genes. This method is preferred to test the effect of PG1 expression in animal that are spontaneously developing tumor. The production of specific adenoviruses is obtained using methods familiar to those with normal skills in cell and molecular biology.

II. POLYNUCLEOTIDES:

The present invention encompasses polynucleotides in the form of PG1 genomic or cDNA as well as polynucleotides for use as primers and probes in the methods of the invention. These polynucleotides may consist of, consist essentially of, or comprise a contiguous span of nucleotides of 10 a sequence from any sequence in the Sequence Listing as well as sequences which are complementary thereto ("complements thereof"). Preferably said sequence is selected from SEQ ID NOs: 3, 112-125, 179, 182-184. The "contiguous span" is at least 6, 8, 10, 12, 15, 20, 25, 30, 50, 100, 200, or 500 nucleotides in length. It should be noted that the polynucleotides of the present invention are not limited to having the exact flanking sequences surrounding the polymorphic bases which are enumerated in Sequence Listing. Rather, it will be appreciated that the flanking sequences surrounding the biallelic markers, or any of the primers of probes of the invention which are more distant from a biallelic markers, is lengthened or shortened to any extent compatible with their intended use and the present invention specifically contemplates such sequences. It will be appreciated that the polynucleotides referred to in the Sequence Listing is of any length compatible 20 with their intended use. Also the flanking regions outside of the contiguous span need not be homologous to native flanking sequences which actually occur in humans. The addition of any nucleotide sequence, which is compatible with the nucleotides intended use is specifically contemplated. The contiguous span may optionally include the PG1-related biallelic marker in said sequence. Optionally either allele of the biallelic markers described above in the definition of PG1-25 related biallelic marker is specified as being present at the PG1-related biallelic marker.

The invention also relates to polynucleotides that hybridize, under conditions of high or intermediate stringency, to a polynucleotide of a sequence from any sequence in the Sequence Listing as well as sequences, which are complementary thereto. Preferably said sequence is selected from SEQ ID NOs: 3, 112-125, 179, 182-184. Preferably such polynucleotides is at least 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 200, or 500 nucleotides in length. Preferred polynucleotides comprise an PG1-related biallelic marker. Optionally either allele of the biallelic markers described above in the definition of PG1-related biallelic marker is specified as being present at the biallelic marker site. Conditions of high and intermediate stringency are further described in Section X.C.4, below.

35 The invention embodies polynucleotides which encode an entire human, mouse or mammalian PG1 protein, or fragments thereof. Generally the polynucleotides of the invention

comprise the naturally occurring nucleotide sequence of the PG1. However, any naturally occurring silent codon variation or other silent codon variation can be employed to encode the PG1 amino acids sequence. As for those amino acids which are changed or added to the PG1 gene for any embodiment of the invention which requires the expression of a nucleotide sequence, the nucleic acid sequences 5 generally will be chosen to optimize expression in the specific human or non-human animal system in which the polynucleotide is intended to be used, making use of known codon preferences. The PG1 polynucleotides of the invention can be the native nucleotide sequence which encodes a human, mouse, or mammalian PG1 protein, preferably the PG1 polynucleotide sequence of SEQ ID NOs: 3, 112-125, 179, 182-184, and the compliments thereof. The polynucleotides of the invention include 10 those which encode PG1 polypeptides with a contiguous stretch of at least 8, 10, 12, 15, 20, 25, 30, 50, 100 or 200 amino acids from SEQ ID NOs: 4, 5, 70, 74, and 125-136, as well as any other human, In addition the present invention encompasses mouse or mammalian PG1 polypeptide. polynucleotides which comprise a contiguous stretch of at least 8, 10, 12, 15, 20, 25, 30, 50, 100, 200, 500 nucleotides of a human, mouse or mammalian PG1 genomic sequence as well as complete human, mouse, or mammalian PG1 genes, preferably of SEQ ID NOs: 179, 182, 183, and the compliments 15 thereof.

The present invention encompasses polynucleotides which consist of, consist essentially of, or comprise a contiguous stretch of at least 8, 10, 12, 15, 20, 25, 30, 50, 100, 200, or 500 nucleotides of a human, mouse or mammalian PG1 cDNA sequences as well as an entire human, mouse, or 20 mammalian PG1 cDNA. The cDNA species and polynucleotide fragments comprised by the polynucleotides of the invention include the predominant species derived from any human, mouse or mammal source, preferably SEQ ID NOs: 3, 184, and the compliments thereof. In addition, the polynucleotides of the invention comprise cDNA species, and fragments thereof, that result from the alternative splicing of PG1 transcripts in any human, mouse or other mammal, preferably the cDNA species of SEQ ID NOs: 112-124, and compliments thereof. Moreover, the invention encompasses cDNA species and other polynucleotides which consist of or comprise the polynucleotides which span a splice junction, preferably including any one of SEQ ID NOs: 137 to 178, and the compliments thereof; more preferably any one of SEQ ID NOs: 137 to 149, 151 to 169, 171 to 178, and the compliments thereof. The polynucleotides of the invention also include cDNA and other polynucleotides which comprise two covalently linked PG1 exons, derived from a single human, mouse or mammalian species, immediately adjacent to one another in the order shown, and selected from the following pairs of PG1 exons: 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 2:3, 2:4, 2:5, 2:6, 2:7, 2:8, 3:4, 3:5, 3:6, 3:7, 3:8, 4:5, 4:6, 4:7, 4:8, 5:6, 5:7, 5:8, 6:7, 6:8, 7:8, 1:1bis, 1bis:2, 1bis:3, 1bis:4, 1bis:5, 1bis:6, 1bis:7, 1bis:8, 3:3bis, 3bis:4, 3bis:5, 3bis:6, 3bis:7, 3bis:8, 5:5bis, 5bis:6, 5bis:7, 5bis:8, 35 1:6bis, 2:6bis, 3:6bis, 4:6bis, 5:6bis, 6bi., 6bis, embodiment the sequences of the PG1 exons in each of the pairs of exons is selected as follows:

```
exon 1 – SEQ ID NO: 100; exon 2 – SEQ ID NO: 101; exon 3 – SEQ ID NO: 102;
exon 4 - SEQ ID NO: 103; exon 5 - SEQ ID NO: 104; exon 6 - SEQ ID NO: 105;
exon 7 - SEQ ID NO: 106; exon 8 - SEQ ID NO: 107; exon 1bis - SEQ ID NO: 108;
exon 3bis - SEQ ID NO: 109; exon 5bis - SEQ ID NO: 110; and
```

5 exon 6bis - SEQ ID NO: 111. Because of the 8 different polyadenylation sites in exon 8, any cDNA or polynucleotide of the invention comprising a human cDNA fragment encompassing exon 8 is truncated such that only the first 330 nucleotides, 699 nucleotides, 833 nucleotides, 1826 nucleotides, 2485 nucleotides, 2805 nucleotides, 4269 nucleotides or 4315 nucleotides of exon 8 shown in SEQ ID NO: 107 are present.

The primers of the present invention is designed from the disclosed sequences for any method known in the art. A preferred set of primers is fashioned such that the 3' end of the contiguous span of identity with the sequences of the Sequence Listing is present at the 3' end of the primer. Such a configuration allows the 3' end of the primer to hybridize to a selected nucleic acid sequence and dramatically increases the efficiency of the primer for amplification or sequencing reactions. Allele 15 specific primers is designed such that a biallelic marker is at the 3' end of the contiguous span and the contiguous span is present at the 3' end of the primer. Such allele specific primers tend to selectively prime an amplification or sequencing reaction so long as they are used with a nucleic acid sample that contains one of the two alleles present at a biallelic marker. The 3' end of primer of the invention is located within or at least 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500, or 1000 nucleotides 20 upstream of an PG1-related biallelic marker in said sequence or at any other location which is appropriate for their intended use in sequencing, amplification or the location of novel sequences or markers.

Preferred amplification primers include the polynucleotides disclosed in SEQ ID NOs: 39-56, and 63-68. Additional preferred amplification primers for particular non-genic PG1-related biallelic markers are listed as follows by the internal reference number for the marker and the SEQ ID NOs for the PU and RP amplification primers respectively:

```
4-14-107 use SEQ ID NOs 339 and 382; 4-14-317 use SEQ ID NOs 339 and 382;
4-14-35 use SEQ ID NOs 339 and 382; 4-20-149 use SEQ ID NOs 340 and 383;
4-22-174 use SEQ ID NOs 341 and 384; 4-22-176 use SEQ ID NOs 341 and 384;
```

30 4-26-60 use SEQ ID NOs 342 and 385; 4-26-72 use SEQ ID NOs 342 and 385;

4-3-130 use SEQ ID NOs 343 and 386; 4-38-63 use SEQ ID NOs 344 and 387;

4-38-83 use SEQ ID NOs 344 and 387; 4-4-152 use SEQ ID NOs 345 and 388;

4-4-187 use

SEQ ID NOs 345 and 388; 4-4-288 use SEQ ID NOs 345 and 388; 4-42-304 use SEQ ID NOs 346 and 389; 4-42-401 use SEQ ID NOs 346 and 389;

35 4-13-328 use SEQ ID NOs 347 and 390; 4-43-70 use SEQ ID NOs 347 and 390;

4-50-209 use SEQ ID NOs 348 and 391; 4-50-293 use SEQ ID NOs 348 and 391;

```
4-50-323 use SEQ ID NOs 348 and 391; 4-50-329 use SEQ ID NOs 348 and 391;
    4-50-330 use SEQ ID NOs 348 and 391; 4-52-163 use SEQ ID NOs 349 and 392;
    4-52-88 use SEQ ID NOs 349 and 392; 4-53-258 use SEQ ID NOs 350 and 393;
    4-54-283 use SEQ ID NOs 351 and 394; 4-54-388 use SEQ ID NOs 351 and 394;
5 4-55-70 use SEQ ID NOs 352 and 395; 4-55-95 use SEQ ID NOs 352 and 395;
    4-56-159 use SEQ ID NOs 353 and 396; 4-56-213 use SEQ ID NOs 353 and 396;
    4-58-289 use SEQ ID NOs 354 and 397; 4-58-318 use SEQ ID NOs 354 and 397;
    4-60-266 use SEQ ID NOs 355 and 398; 4-60-293 use SEQ ID NOs 355 and 398;
    4-84-241 use SEQ ID NOs 356 and 399; 4-84-262 use SEQ ID NOs 356 and 399;
10 4-86-206 use SEQ ID NOs 357 and 400; 4-86-309 use SEQ ID NOs 357 and 400;
    4-88-349 use SEQ ID NOs 358 and 401; 4-89-87 use SEQ ID NOs 359 and 402;
    99-123-184 use SEQ ID NOs 360 and 403; 99-128-202 use SEQ ID NOs 361 and 404;
                                                                                         99-128-
     275 use SEQ ID NOs 361 and 404; 99-128-313 use SEQ ID NOs 361 and 404;
     99-128-60 use SEQ ID NOs 361 and 404; 99-12907-295 use SEQ ID NOs 362 and 405;
                                                                                             99-
15 130-58 use SEQ ID NOs 363 and 406; 99-134-362 use SEQ ID NOs 364 and 407;
                                                                                     99-140-130
     use SEQ ID NOs 365 and 408; 99-1462-238 use SEQ ID NOs 366 and 409;
                                                                                  99-147-181 use
     SEQ ID NOs 367 and 410; 99-1474-156 use SEQ ID NOs 368 and 411;
                                                                            99-1474-359 use SEQ
     ID NOs 368 and 411;
     99-1479-158 use SEQ ID NOs 369 and 412;
20 99-1479-379 use SEQ ID NOs 369 and 412; 99-148-129 use SEQ ID NOs 370 and 413;
     99-148-132 use SEQ ID NOs 370 and 413; 99-148-139 use SEQ ID NOs 370 and 413;
     99-148-140 use SEQ ID NOs 370 and 413; 99-148-182 use SEQ ID NOs 370 and 413;
     99-148-366 use SEQ ID NOs 370 and 413; 99-148-76 use SEQ ID NOs 370 and 413;
                                                                                        99-1480-
     290 use SEQ ID NOs 371 and 414;
 25 99-1481-285 use SEQ ID NOs 372 and 415;
      99-1484-101 use SEQ ID NOs 373 and 416;
      99-1484-328 use SEQ ID NOs 373 and 416;
      99-1485-251 use SEQ ID NOs 374 and 417;
      99-1490-381 use SEQ ID NOs 375 and 418;
 30 99-1493-280 use SEQ ID NOs 376 and 419; 99-151-94 use SEQ ID NOs 377 and 420;
      99-211-291 use SEQ ID NOs 378 and 421; 99-213-37 use SEQ ID NOs 379 and 422;
      99-221-442 use SEQ ID NOs 380 and 423; 99-222-109 use SEQ ID NOs 381 and 424; and the
      compliments thereof.
              Primers with their 3' ends located 1 nucleotide upstream or downstream of a PG1-related
```

Primers with their 3' ends located 1 nucleotide upstream or downstream of a PG1-related
35 biallelic marker have a special utility in microsequencing assays. Preferred microsequencing primers
include the polynucleotides from position 1 to position 23 and from position 25 to position 47 of SEQ

ID NOs: 21-38, and as well as the compliments thereof. Additional preferred microsequencing primers for particular non-genic PG1-related biallelic markers are listed as follows by the internal reference number for the marker and the SEQ ID NOs of the two preferred microsequencing primers: 4-14-107 of SEQ ID NOs 425 and 502*; 4-14-317 of SEQ ID NOs 426 and 503*;

- 5 4-14-35 of SEQ ID NOs 427 and 504*; 4-20-149 of SEQ ID NOs 428* and 505; 4-20-77 of SEQ ID NOs 429 and 506; 4-22-174 of SEQ ID NOs 430* and 507; 4-22-176 of SEQ ID NOs 431 and 508; 4-26-60 of SEQ ID NOs 432 and 509*; 4-26-72 of SEQ ID NOs 433 and 510; 4-3-130 of SEQ ID NOs 434 and 511*; 4-38-63 of SEQ ID NOs 435 and 512; 4-38-83 of SEQ ID NOs 436 and 513*;
- 4-4-152 of SEQ ID NOs 437 and 514; 4-4-187 of SEQ ID NOs 438* and 515;
 4-4-288 of SEQ ID NOs 439 and 516; 4-42-304 of SEQ ID NOs 440 and 517;
 4-42-401 of SEQ ID NOs 441* and 518; 4-43-328 of SEQ ID NOs 442 and 519;
 4-43-70 of SEQ ID NOs 443* and 520; 4-50-209 of SEQ ID NOs 444* and 521;
 4-50-293 of SEQ ID NOs 445* and 522; 4-50-323 of SEQ ID NOs 446* and 523;
- 4-50-329 of SEQ ID NOs 447* and 524; 4-50-330 of SEQ ID NOs 448 and 525;
 4-52-163 of SEQ ID NOs 449* and 526; 4-52-88 of SEQ ID NOs 450* and 527;
 4-53-258 of SEQ ID NOs 451 and 528*;4-54-283 of SEQ ID NOs 452* and 529;
 4-54-388 of SEQ ID NOs 453 and 530; 4-55-70 of SEQ ID NOs 454 and 531*;
 4-55-95 of SEQ ID NOs 455* and 532; 4-56-159 of SEQ ID NOs 456* and 533;
- 4-56-213 of SEQ ID NOs 457 and 534; 4-58-289 of SEQ ID NOs 458* and 535;
 4-58-318 of SEQ ID NOs 459* and 536; 4-60-266 of SEQ ID NOs 460* and 537;
 4-60-293 of SEQ ID NOs 461* and 538; 4-84-241 of SEQ ID NOs 462 and 539*;
 4-84-262 of SEQ ID NOs 463 and 540; 4-86-206 of SEQ ID NOs 464 and 541*;
 4-86-309 of SEQ ID NOs 465 and 542; 4-88-349 of SEQ ID NOs 466 and 543.;
- 25 4-89-87 of SEQ ID NOs 467* and 544.; 99-123-184 of SEQ ID NOs 468 and 545; 99-128-202 of SEQ ID NOs 469 and 546; 99-128-275 of SEQ ID NOs 470 and 547; 99-128-313 of SEQ ID NOs 471 and 548; 99-128-60 of SEQ ID NOs 472* and 549; 99-12907-295 of SEQ ID NOs 473 and 550*; 99-130-58 of SEQ ID NOs 474* and 551*;
- 30 99-134-362 of SEQ ID NOs 475 and 552*; 99-140-130 of SEQ ID NOs 476* and 553*; 99-1462-238 of SEQ ID NOs 477* and 554; 99-147-181 of SEQ ID NOs 478 and 555*; 99-1474-156 of SEQ ID NOs 479 and 556*; 99-1474-359 of SEQ ID NOs 480 and 557; 99-1479-158 of SEQ ID NOs 481* and 558; 99-1479-379 of SEQ ID NOs 482 and 559; 99-148-129 of SEQ ID NOs 483 and 560; 99-148-132 of SEQ ID NOs 484 and 561;
- 99-148-139 of SEQ ID NOs 485 and 562; 99-148-140 of SEQ ID NOs 486 and 563;99-148-182 of SEQ ID NOs 487 and 564*; 99-148-366 or SEQ ID NOs 488 and 565;

```
99-148-76 of SEQ ID NOs 489 and 566; 99-1480-290 of SEQ ID NOs 490 and 567*; 99-1481-285 of SEQ ID NOs 491 and 568*; 99-1484-101 of SEQ ID NOs 492 and 569; 99-1484-328 of SEQ ID NOs 493* and 570; 99-1485-251 of SEQ ID NOs 494 and 571*; 599-1490-381 of SEQ ID NOs 495* and 572; 99-1493-280 of SEQ ID NOs 496 and 573*; 99-151-94 of SEQ ID NOs 497 and 574*; 99-211-291 of SEQ ID NOs 498* and 575; 99-213-37 of SEQ ID NOs 499 and 576; 99-221-442 of SEQ ID 500 and 577; 99-222-109 of SEQ ID NOs 501* and 578; and compliments thereof.
```

Additional preferred microsequencing primers for particular genic PG1-related biallelic markers include a polynucleotide selected from the group consisting of the nucleotide sequences from position N-X to position N-1 of SEQ ID NO:179, nucleotide sequences from position N+X of SEQ ID NO:179, and the compliments thereof, wherein X is equal to 15, 18, 20, 25, 30, or a range of 15 to 30, and N is equal to one of the following values: 2159; 2443; 4452; 5733; 8438; 11843; 1983; 12080; 12221; 12947; 13147; 13194; 13310; 13342; 13367; 13594; 13680; 13902; 16231; 16388; 17608; 18034; 18290; 18786; 22835; 22872; 25183; 25192; 25614; 26911; 32703; 34491; 34756; 34934; 5160; 39897; 40598; 40816; 40947; 45783; 47929; 48206; 48207; 49282; 50037; 50054; 50101; 50220; 50440; 50562; 50653; 50660; 50745; 50885; 51249; 51333; 51435; 51468; 51515; 51557; 51566; 51632; 51666; 52016; 52096; 52151; 52282; 52348; 52410; 52580; 52712; 52772; 52860; 53092; 53272; 53389; 53511; 53600; 53665; 53815; 54365; and 54541.

The probes of the present invention is designed from the disclosed sequences for any method known in the art, particularly methods which allow for testing if a particular sequence or marker disclosed herein is present. A preferred set of probes is designed for use in the hybridization assays of the invention in any manner known in the art such that they selectively bind to one allele of a biallelic marker, but not the other under any particular set of assay conditions. Preferred hybridization probes may consists of, consist essentially of, or comprise a contiguous span which ranges in length from 8, 10, 12, 15, 18 or 20 to 25, 35, 40, 50, 60, 70, or 80 nucleotides, or be specified as being 12, 15, 18, 20, 25, 35, 40, or 50 nucleotides in length and including a PG1-related biallelic marker of said sequence. Optionally either of the two alleles specified in the definition of PG1-realted biallelic marker is 30 specified as being present at the biallelic marker site. Optionally, said biallelic marker is within 6, 5, 4, 3, 2, or 1 nucleotides of the center of the hybridization probe or at the center of said probe. A preferred set of hybridization probes is disclosed in SEQ ID NOs: 21-38, 57-62, 185-338, and the compliments thereof. Another particularly preferred set of hybridization probes includes the polynucleotides from position X to position Y of any one of SEQ ID NOs: 21-38, 57-62, 185-338, or 35 the compliments thereof, wherein X is equal to 5, 8, 10, 12, 14, 16, 18 or a range of 5 to 18, and Y is equal to 30, 32, 34, 36, 38, 40, 43 or a range of 30 to 43; preferably X equals 12 and Y equals 36. WO 99/32644 PCT/IB98/02133

Additional preferred hybridization probes for particular genic PG1-related biallelic markers include a polynucleotide selected from the group consisting of the nucleotide sequences from position N-X to position N+Y of SEQ ID NO:179, and the compliments thereof, wherein X is equal to 8, 10, 12, 15, 20, 25, or a range of 8 to 30, Y is equal to 8, 10, 12, 15, 20, 25, or a range of 8 to 30, and N is equal to one of the following values: 2159; 2443; 4452; 5733; 8438; 11843; 1983; 12080; 12221; 12947; 13147; 13194; 13310; 13342; 13367; 13594; 13680; 13902; 16231; 16388; 17608; 18034; 18290; 18786; 22835; 22872; 25183; 25192; 25614; 26911; 32703; 34491; 34756; 34934; 5160; 39897; 40598; 40816; 40947; 45783; 47929; 48206; 48207; 49282; 50037; 50054; 50101; 50220; 50440; 50562; 50653; 50660; 50745; 50885; 51249; 51333; 51435; 51468; 51515; 51557; 51566; 51632; 51666; 52016; 52096; 52151; 52282; 52348; 52410; 52580; 52712; 52772; 52860; 53092; 53272; 53389; 53511; 53600; 53665; 53815; 54365; and 54541; wherein the nucleotide at position N is selected from one of the two alleles specified in the definition of PG1-realted biallelic marker at the biallelic marker site at position N.

Any of the polynucleotides of the present invention can be labeled, if desired, by 15 incorporating a label detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive substances, fluorescent dyes or biotin. Preferably, polynucleotides are labeled at their 3' and 5' ends. A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or 20 probes and can be a specific binding member which forms a binding pair with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending upon the type of label carried by a polynucleotide or a probe, it is employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it is selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase. DNA Labeling techniques are well known to the skilled technician.

Any of the polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes® and others. The solid support is not critical and can be selected by one skilled in the art. Thus, latex

particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to 10 the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes® and other configurations known to those of ordinary skill in the art. The polynucleotides of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the inventions to a single solid support. In addition, polynucleotides other than those of the invention may attached to the same solid support as one or more polynucleotides of the invention.

Any polynucleotide provided herein is attached in overlapping areas or at random locations on the solid support. Alternatively the polynucleotides of the invention is attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these "addressable" arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the GenechipsTM, and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092. These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods, which in orporate a combination of photolithographic methods and solid phase oligonucleotide synthesis (Fodor et al., Science, 251:767-777, 1991). The immobilization of arrays of oligonucleotides on solid

WO 99/32644 PCT/IB98/02133

supports has been rendered possible by the development of a technology generally identified as "Very Large Scale Immobilized Polymer Synthesis" (VLSIPSTM) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPSTM technologies are provided in US Patents 5,143,854 and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques. In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and sequence information. Examples of such presentation strategies are disclosed in PCT Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256.

Oligonucleotide arrays may comprise at least one of the sequences selected from the group consisting of SEQ ID NOs: 3, 21-38, 57-62, 100-124, 179, 185-338, the preferred hybridization probes for genic PG1-related biallelic markers described above; and the sequences complementary thereto; or a fragment thereof of at least 15 consecutive nucleotides for determining whether a sample contains one or more alleles of the biallelic markers of the present invention. Oligonucleotide arrays may also comprise at least one of the sequences selected from the group consisting of SEQ ID NOs: 179, 339-424; and the sequences complementary thereto or a fragment thereof of at least 15 consecutive nucleotides for amplifying one or more alleles of the PG1-realted biallelic markers. In other embodiments, arrays may also comprise at least one of the sequences selected from the group consisting of SEQ ID 425-578, the preferred microsequencing primers for genic PG1-related biallelic markers described above; and the sequences complementary thereto or a fragment thereof of at least 15 consecutive nucleotides for conducting microsequencing analyses to determine whether a sample contains one or more alleles of PG1-related biallelic marker.

The present invention further encompasses polynucleotide sequences that hybridize to any one of SEQ ID NOs: 3, 69, 100-112, or 179-184 under conditions of high or intermediate stringency as described below:

(i) By way of example and not limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 μg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C, the preferred hybridization temperature, in prehybridization mixture containing 100 μg/ml denatured salmon sperm DNA and 5-20 X 10⁶ cpm of ³²P-labeled probe. Alternatively, the hybridization step can be performed at 65°C in the presence of SSC buffer, 1 x SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1X SSC at 50°C for 45 min. Alternatively, filter washes can be performed in a solution containing 2 x SSC and 0.1% SDS,

or 0.5 x SSC and 0.1% SDS, or 0.1 x SSC and 0.1% SDS at 68°C for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. Other conditions of high stringency which is used are well known in the art and as cited in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., pp. 9.47-9.57; and 5 Ausubel et al., 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. Preferably, such sequences encode a homolog of a polypeptide encoded by one of ORF2 to ORF1297. In one embodiment, such sequences encode a mammalian PG1 polypeptide.

(ii) By way of example and not limitation, procedures using conditions of intermediate stringency are as follows: Filters containing DNA are prehybridized, and then hybridized at a 10 temperature of 60°C in the presence of a 5 x SSC buffer and labeled probe. Subsequently, filters washes are performed in a solution containing 2x SSC at 50°C and the hybridized probes are detectable by autoradiography. Other conditions of intermediate stringency which is used are well known in the art and as cited in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., pp. 9.47-9.57; and Ausubel et al., 1989, Current 15 Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. Preferably, such sequences encode a homolog of a polypeptide encoded by one of SEQ ID NOs: 3, 69, 100-112, or 179-184. In one embodiment, such sequences encode a mammalian PG1 polypeptide.

The present invention also encompasses diagnostic kits comprising one or more polynucleotides of the invention with a portion or all of the necessary reagents and instructions for genotyping a test subject by determining the identity of a nucleotide at a PG1-related biallelic marker. The polynucleotides of a kit may optionally be attached to a solid support, or be part of an array or addressable array of polynucleotides. The kit may provide for the determination of the identity of the nucleotide at a marker position by any method known in the art including, but not limited to, a sequencing assay method, a microsequencing assay method, a hybridization assay method, or an allele 25 specific amplification method. Optionally such a kit may include instructions for scoring the results of the determination with respect to the test subjects' risk of contracting a cancer or prostate cancer, or likely response to an anti-cancer agent or anti-prostate cancer agent, or chances of suffering from side effects to an anti-cancer agent or anti-prostate cancer agent.

Use of PG1 Nucleic Acids as Reagents

The PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, 112-124 and 30 PG1 alleles responsible for a detectable phenotype (such as those obtainable by the methods of Example 12, and SEQ ID NO:69) can be used to prepare PCR primers for use in diagnostic techniques or genetic engineering methods such as those described above. Example 10 describes the use of the PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, 112-124 and PG1 alleles responsible for a detectable phenotype (such as those obtainable by the methods of Example 12) in PCR amplification procedures.

Example 10

The PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, and PG1 alleles responsible for a detectable phenotype (such as those obtainable by the methods of Example 12) is used to prepare PCR primers for a variety of applications, including isolation procedures for cloning nucleic acids capable of hybridizing to such sequences, diagnostic techniques and forensic techniques. The PCR primers comprise at least 10 consecutive bases of the PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, 112-124 and PG1 alleles responsible for a detectable phenotype (such as those obtainable by the methods of Example 12) or the sequences complementary thereto. Preferably, the PCR primers comprise at least 12, 15, or 17 consecutive bases of these sequences. More preferably, the PCR primers comprise at least 20-30 consecutive bases of the PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, 112-124 and PG1 alleles responsible for a detectable phenotype (such as those obtainable by the methods of Example 12) or the sequences complementary thereto. In some embodiments, the PCR primers may comprise more than 30 consecutive bases of the PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, 112-124 and PG1 alleles responsible for a 15 detectable phenotype (such as those obtainable by the methods of Example 12) or the sequences complementary thereto. It is preferred that the primer pairs to be used together in a PCR amplification have approximately the same G/C ratio, so that melting temperatures are approximately the same.

A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see Molecular Cloning to Genetic Engineering White, B.A. Ed. in Methods in Molecular Biology 67: Humana Press, Totowa 1997. In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs; and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended.

Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites.

The polynucleotides of the Invention also encompass vectors and DNA constructs as well as other forms of primers and probes. For a thorough description of these embodiments please see 30 Sections VIII, X, and XI below.

III. POLYPEPTIDES

PG1 Proteins and Polypeptide Fragments

The term "PG1 polypeptides" is used herein to embrace all of the proteins and polypeptides of the present invention. Also forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. The invention embodies PG1 proteins from human (SEQ ID NOs: 4, and 5), and mouse (SEQ ID NO: 74).

However, PG1 species from other varieties of mammals are expressly contemplated and is isolated using the antibodies of the present invention in conjunction with standard affinity chromatography methods as well as being expressed from the PG1 genes isolated from other mammalian sources using human and mouse PG1 nucleic acid sequences as primers and probes as well as the methods described herein.

The invention also embodies PG1 proteins translated from less common alternative splice species, including SEQ ID NOs: 125-136, and PG1 proteins which result from naturally occurring mutant, particularly functional mutants of PG1, including SEQ ID NO: 70, which is identified and obtained by the described herein. The present invention also embodies polypeptides comprising a contiguous stretch of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 50, or 100 amino acids of a PG1 protein. In a preferred embodiment the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the PG1 protein sequence. For instance, polypeptides that contain either the Arg and His residues at amino acid position 184, and polypeptides 15 that contain either the Arg or Ile residue at amino acid position 293 of the SEQ ID NO: 4 in said contiguous stretch are particularly preferred embodiments of the invention and useful in the manufacture of antibodies to detect the presence and absence of these mutations. Similarly, polypeptides with a carboxy terminus at position 228 is a particularly preferred embodiment of the invention and useful in the manufacture of antibodies to detect the presence and absence of the mutation shown in SEQ ID NOs: 69 and 70.

Similarly, polypeptides that that contain an peptide sequences of 8, 10, 12, 15, or 25 amino acids encoded over a naturally-occurring splice junction (the point at which two human PG1 exon (SEQ ID NOs: 100-111) are covalently linked) in said contiguous stretch are particularly preferred embodiments and useful in the manufacture of antibodies to detect the presence, localization, and quantity of the various protein products of the PG1 alternative splice species.

PG1 proteins are preferably isolated from human, mouse or mammalian tissue samples or expressed from human, mouse or mammalian genes.

The PG1 polypeptides of the invention can be made using routine expression methods known in the art, see, for instance, Example 11, below. The polynucleotide encoding the desired polypeptide, is ligated into an expression vector suitable for any convenient host. Both eukaryotic and prokaryotic host systems is used in forming recombinant polypeptides, and a summary of some of the more common systems are included in Sections II and VIII. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification is by any technique known in the art, for example, differential extraction, salt fractionation, chromatography, centrifugation, and the like. See, for example, Methods in Enzymology for L variety of methods for purifying proteins.

5

In addition, shorter protein fragments is produced by chemical synthesis. Alternatively the proteins of the invention is extracted from cells or tissues of humans or non-human animals. Methods for purifying proteins are known in the art, and include the use of detergents or chaotropic agents to disrupt particles followed by differential extraction and separation of the polypeptides by ion exchange chromatography, affinity chromatography, sedimentation according to density, and gel electrophoresis.

Expression of the PG1 Protein

Any PG1 cDNA, including SEQ ID NO: 3, 69, 112-124, or 184 or synthetic DNAs is use as described in Example 11 below to express PG1 proteins and polypeptides.

10 Example 11

The nucleic acid encoding the PG1 protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The PG1 insert in the expression vector may comprise the full coding sequence for the PG1 protein or a portion thereof. For example, the PG1 derived insert may encode a polypeptide comprising at least 10 consecutive amino acids of the PG1 proteins of SEQ ID NO: 4.

The expression vector is any of the mammalian yeast, insect or bacterial expression systems known in the art, see for example Section VIII. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence is optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, et al., U.S. Patent No. 5,082,767.

The following is provided as one exemplary method to express the PG1 protein or a portion thereof. In one embodiment, the entire coding sequence of the PG1 cDNA through the poly A signal of the cDNA are operably linked to a promoter in the expression vector. Alternatively, if the nucleic acid encoding a portion of the PG1 protein lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the insert from the PG1 cDNA lacks a poly A signal, this sequence can be added to the construct by, for example, splicing out the Poly A signal from pSG5 (Stratagene) using BgII and SaII restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the gag gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex Thymidine Kinase promoter and the selectable neomycin gene. The nucleic acid encoding the PG1 protein or a portion thereof is obtained by PCR from a bacterial vector containing the PG1 cDNA of SEQ ID NO: 3 using oligonacleotide primers complementary to the PG1 cDNA or portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5• primer and

PCT/IB98/02133 WO 99/32644

BglII at the 5• end of the corresponding cDNA 3• primer, taking care to ensure that the sequence encoding the PG1 protein or a portion thereof is positioned properly with respect to the poly A signal. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1, now containing a poly A signal and 5 digested with BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600ug/ml G418 (Sigma, St. Louis, Missouri).

Alternatively, the nucleic acids encoding the PG1 protein or a portion thereof is cloned into pED6dpc2 (Genetics Institute, Cambridge, MA). The resulting pED6dpc2 constructs is transfected into a suitable host cell, such as COS 1 cells. Methotrexate resistant cells are selected and expanded.

The above procedures may also be used to express a mutant PG1 protein responsible for a detectable phenotype or a portion thereof.

The expressed proteins is purified using conventional purification techniques such as ammonium sulfate precipitation or chromatographic separation based on size or charge. The protein encoded by the nucleic acid insert may also be purified using standard immunochromatography techniques. In such procedures, a solution containing the expressed PG1 protein or portion thereof, such as a cell extract, is applied to a column having antibodies against the PG1 protein or portion thereof is attached to the 20 chromatography matrix. The expressed protein is allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound expressed protein is then released from the column and recovered using standard techniques.

To confirm expression of the PG1 protein or a portion thereof, the proteins expressed from host cells containing an expression vector containing an insert encoding the PG1 protein or a portion thereof can be compared to the proteins expressed in host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the PG1 protein or a portion thereof is being expressed. Generally, the band will have the mobility expected for the PG1 protein or portion thereof. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Antibodies capable of specifically recognizing the expressed PG1 protein or a portion thereof is generated as described below in Section VII.

If antibody production is not possible, the nucleic acids encoding the PG1 protein or a portion thereof is incorporated into expression vectors designed for use in purification schemes employing 35 chimeric polypeptides. In such strategies the nucleic acid encoding the PG1 protein or a portion thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera

10

is β -globin or a nickel binding polypeptide encoding sequence. A chromatography matrix having antibody to β -globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites is engineered between the β -globin gene or the nickel binding polypeptide and the PG1 protein or portion thereof. Thus, the two polypeptides of the chimera is separated from one another by protease digestion.

One useful expression vector for generating β-globin chimerics is pSG5 (Stratagene), which encodes rabbit β-globin. Intron II of the rabbit β-globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis et al., (Basic Methods in Molecular Biology, L.G. Davis, M.D. Dibner, and J.F. Battey, ed., Elsevier Press, NY, 1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using in vitro translation systems such as the In vitro ExpressTM Translation Kit (Stratagene).

IV. IDENTIFICATION OF MUTATIONS IN THE PG1 GENE WHICH ARE ASSOCIATED WITH A DETECTABLE PHENOTYPE

Mutations in the PG1 gene which are responsible for a detectable phenotype is identified by comparing the sequences of the PG1 genes from affected and unaffected individuals as described in Example 12, below. The detectable phenotype may comprise a variety of manifestations of altered PG1 function, including prostate cancer, hepatocellular carcinoma, colorectal cancer, non-small cell lung cancer, squamous cell carcinoma, or other conditions. The mutations may comprise point mutations, deletions, or insertions of the PG1 gene. The mutations may lie within the coding sequence for the PG1 protein or within regulatory regions in the PG1 gene.

Example 12

Oligonucleotide primers are designed to amplify the sequences of each of the exons or the promoter region of the PG1 gene. The oligonucleotide primers may comprise at least 10 consecutive nucleotides of the PG1 genomic DNA of SEQ ID NO:179 or the PG1 cDNA of SEQ ID NO: 3 or the sequences complementary thereto. Preferably, the oligonucleotides comprise at least 15 consecutive nucleotides of the PG1 genomic DNA of SEQ ID NO:179 or the PG1 cDNA of SEQ ID NO: 3 or the sequences complementary thereto. In some embodiments, the oligonucleotides may comprise at least 20 consecutive nucleotides of the PG1 genomic DNA of SEQ ID NO: 179 or the PG1 cDNA of SEQ ID NO:3 or the sequences complementary thereto. In other embodiments, the oligonucleotides may comprise 25 or more consecutive nucleotides of the PG1 genomic DNA of SEQ ID NO: 179 or the PG1 cDNA of SEQ ID NO: 3 or the sequences complementary thereto.

Each primer pair is used to amplify the exon or promoter region from which it is derived.

35 Amplification is carried out on genomic DNA samples from affected patients and unaffected controls using the PCR conditions described above. Amplification products from the genomic PCRs are

subjected to automated dideoxy terminator sequencing reactions and electrophoresed on ABI 377 sequencers. Following gel image analysis and DNA sequence extraction, ABI sequence data are automatically analyzed to detect the presence of sequence variations among affected and unaffected individuals. Sequences are verified by determining the sequences of both DNA strands for each individual. Preferably, these candidate mutations are detected by comparing individuals homozygous for haplotype 5 of Figure 4 and controls not carrying haplotype 5 or related haplotypes.

Candidate polymorphisms suspected of being responsible for the detectable phenotype, such as prostate cancer or other conditions, are then verified by screening a larger population of affected and unaffected individuals using the microsequencing technique described above. Polymorphisms which exhibit a statistically significant correlation with the detectable phenotype are deemed responsible for the detectable phenotype.

Other techniques may also be used to detect polymorphisms associated with a detectable phenotype such as prostate cancer or other conditions. For example, polymorphisms is detected using single stranded conformation analyses such as those described in Orita et al., Proc. Natl. Acad. Sci. U.S.A. 86: 2776-2770 (1989). In this approach, polymorphisms are detected through altered migration on SSCA gels.

Alternatively, polymorphisms is identified using clamped denaturing gel electrophoresis, heteroduplex analysis, chemical mismatch cleavage, and other conventional techniques as described in Sheffield, V.C. et al, Proc. Natl. Acad. Sci. U.S.A 49:699-706 (1991); White, M.B. et al., Genomics 12:301-306 (1992); Grompe, M. et al., Proc. Natl. Acad. Sci. U.S.A 86:5855-5892 (1989); and Grompe, M. Nature Genetics 5:111-117 (1993).

The PG1 genes from individuals carrying PG1 mutations responsible for the detectable phenotype, or cDNAs derived therefrom, is cloned as follows. Nucleic acid samples are obtained from individuals having a PG1 mutation associated with the detectable phenotype. The nucleic acid samples are contacted with a probe derived from the PG1 genomic DNA of SEQ ID NO: 179 or the PG1 cDNA of SEQ ID NO:3. Nucleic acids containing the mutant PG1 allele are identified using conventional techniques. For example, the mutant PG1 gene, or a cDNA derived therefrom, is obtained by conducting an amplification reaction using primers derived from the PG1 genomic DNA of SEQ ID NO: 179 or the PG1 cDNA of SEQ ID NO:3. Alternatively, the mutant PG1 gene, or a cDNA derived therefrom, is identified by hybridizing a genomic library or a cDNA library obtained from an individual having a mutant PG1 gene with a detectable probe derived from the PG1 genomic DNA of SEQ ID NO: 179 or the PG1 cDNA of SEQ ID NO: 3. Alternatively, the mutant PG1 allele is obtained by contacting an expression library from an individual carrying a PG1 mutation with a detectable antibody against the PG1 proteins of SEQ ID NO: 4 or SEQ ID NO: 5 which has been prepared as described below. Those skilled in the art will appreciate that the PG1 genomic DNA of

WO 99/32644 PCT/IB98/02133

SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3 and the PG1 proteins of SEQ ID NOs: 4 and 5 is used in a variety of other conventional techniques to obtain the mutant PG1 gene.

In another embodiment the mutant PG1 allele which causes a detectable phenotype can be isolated by obtaining a nucleic acid sample such as a genomic library or a cDNA library from an individual expressing the detectable phenotype. The nucleic acid sample can be contacted with one or more probes lying in the 8p23 region of the human genome. Nucleic acids in the sample which contain the PG1 gene can be identified by conducting sequencing reactions on the nucleic acids which hybridize to the markers from the 8p23 region of the human genome.

The region of the PG1 gene containing the mutation responsible for the detectable phenotype may also be used in diagnostic techniques such as those described below. For example, oligonucleotides containing the mutation responsible for the detectable phenotype is used in amplification or hybridization based diagnostics, such as those described herein, for detecting individuals suffering from the detectable phenotype or individuals at risk of developing the detectable phenotype at a subsequent time. In addition, the PG1 allele responsible for the detectable phenotype is used in gene therapy as described herein. The PG1 allele responsible for the detectable phenotype may also be cloned into an expression vector to express the mutant PG1 protein a described herein.

During the search for biallelic markers associated with prostate cancer, a number of polymorphic bases were discovered which lie within the PG1 gene. The identities and positions of these polymorphic bases are listed as features in the accompanying Sequence Listing for the PG1 genomic DNA of SEQ ID NO: 179. The polymorphic bases is used in the above-described diagnostic techniques for determining whether an individual is at risk for developing prostate cancer at a subsequent date or suffers from prostate cancer as a result of a PG1 mutation. The identities of the nucleotides present at the polymorphic positions in a nucleic acid sample is determined using the techniques, such as microsequencing analysis, which are described above.

It is possible that one or more of these polymorphisms (or other polymorphic bases) is mutations which are associated with prostate cancer. To determine whether a polymorphism is responsible for prostate cancer, the frequency of each of the alleles in individuals suffering from prostate cancer and unaffected individuals is measured as described in the haplotype analysis above. Those mutations which occur at a statistically significant frequency in the affected population are deemed to be responsible for prostate cancer.

cDNAs containing the identified mutant PG1 gene is prepared as described above and cloned into expression vectors as described below. The proteins expressed from the expression vectors is used to generate antibodies specific for the mutant PG1 proteins as described below. In addition, allele specific probes containing the PG1 mutation responsible for prostate cancer is used in the diagnostic techniques described below.

Genes sharing homology to the PG1 gene is identified as follows.

Example 13

Alternatively, a cDNA library or genomic DNA library to be screened for genes sharing homology to the PG1 gene is obtained from a commercial source or made using techniques familiar to those skilled in the art. The cDNA library or genomic DNA library is hybridized to a detectable probe comprising at least 10 consecutive nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto, using conventional techniques. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto. More preferably, the probe comprises at least 20-30 consecutive nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto. In some embodiments, the probe comprises more than 30 nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto.

Techniques for identifying cDNA clones in a cDNA library which hybridize to a given probe sequence are disclosed in Sambrook et al., Molecular Cloning: A Laboratory Manual 2d Ed., Cold Spring Harbor Laboratory Press, 1989. The same techniques is used to isolate genomic DNAs sharing homology with the PG1 gene.

Briefly, cDNA or genomic DNA clones which hybridize to the detectable probe are identified and isolated for further manipulation as follows. A probe comprising at least 10 consecutive nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto, is labeled with a detectable label such as a radioisotope or a fluorescent molecule. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto. More preferably, the probe comprises 20-30 consecutive nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto. In some embodiments, the probe comprises more than 30 nucleotides from the PG1 cDNA of SEQ ID NO:3, the PG1 genomic DNA of SEQ ID NO: 179, or the sequences complementary thereto.

Techniques for labeling the probe are well known and include phosphorylation with polynucleotide kinase, nick translation, in vitro transcription, and non-radioactive techniques. The cDNAs or genomic DNAs in the library are transferred to a nitrocellulose or nylon filter and denatured.

30 After incubation of the filter with a blocking solution, the filter is contacted with the labeled probe and incubated for a sufficient amount of time for the probe to hybridize to cDNAs or genomic DNAs containing a sequence capable of hybridizing to the probe.

By varying the stringency of the hybridization conditions used to identify cDNAs or genomic DNAs which hybridize to the detectable probe, cDNAs or genomic DNAs having different levels of homology to the probe can be identified and isolated. To identify cDNAs or genomic DNAs having a

WO 99/32644 PCT/IB98/02133

high degree of homology to the probe sequence, the melting temperature of the probe is calculated using the following formulas:

For probes between 14 and 70 nucleotides in length the melting temperature TM is calculated using the formula: Tm=81.5+16.6(log (Na+))+0.41(fraction G+C)-(600/N) where N is the length of the probe.

If the hybridization is carried out in a solution containing formamide, the melting temperature is calculated using the equation Tm=81.5+16.6(log (Na+))+0.41(fraction G+C)-(0.63% formamide)-(600/N) where N is the length of the probe.

Prehybridization is carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100• g denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100• g denatured fragmented salmon sperm DNA, 50% formamide. The formulas for SSC and Denhardt's solutions are listed in Sambrook et al., supra.

Hybridization is conducted by adding the detectable probe to the prehybridization solutions listed above. Where the probe comprises double stranded DNA, it is denatured before addition to the hybridization solution. The filter is contacted with the hybridization solution for a sufficient period of time to allow the probe to hybridize to cDNAs or genomic DNAs containing sequences complementary thereto or homologous thereto. For probes over 200 nucleotides in length, the hybridization is carried out at 15-25. C below the Tm. For shorter probes, such as oligonucleotide probes, the hybridization is conducted at 15-25. C below the Tm. Preferably, for hybridizations in 6X SSC, the hybridization is conducted at approximately 68. C. Preferably, for hybridizations in 50% formamide containing solutions, the hybridization is conducted at approximately 42. C.

All of the foregoing hybridizations would be considered to be under "stringent" conditions.

Following hybridization, the filter is washed in 2X SSC, 0.1% SDS at room temperature for 15 minutes. The filter is then washed with 0.1X SSC, 0.5% SDS at room temperature for 30 minutes to 1 hour. Thereafter, the solution is washed at the hybridization temperature in 0.1X SSC, 0.5% SDS. A final wash is conducted in 0.1X SSC at room temperature.

cDNAs or genomic DNAs homologous to the PG1 gene which have hybridized to the probe are identified by autoradiography or other conventional techniques.

The above procedure is modified to identify cDNAs or genomic DNAs having decreasing levels of homology to the probe sequence. For example, to obtain cDNAs or genomic DNAs of decreasing homology to the detectable probe, less stringent conditions is used. For example, the hybridization temperature is decreased in increments of 5° C from 68° C to 42° C in a hybridization buffer having a Na+ concentration of approximately 1M. Following hybridization, the filter is washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate" conditions above 50° C and "low" conditions below 50° C.

Alternatively, the hybridization is carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42° C. In this case, the concentration of formamide in the hybridization buffer is reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following hybridization, the filter is washed with 6X SSC, 0.5% SDS at 50° C. These conditions are considered to be "moderate" conditions above 25% formamide and "low" conditions below 25% formamide.

cDNAs or genomic DNAs which have hybridized to the probe are identified by autoradiography.

If it is desired to obtain nucleic acids homologous to the PG1 gene, such as allelic variants thereof or nucleic acids encoding proteins related to the PG1 protein, the level of homology between the hybridized nucleic acid and the PG1 gene may readily be determined. To determine the level of homology between the hybridized nucleic acid and the PG1 gene, the nucleotide sequences of the hybridized nucleic acid and the PG1 gene are compared. For example, using the above methods, nucleic acids having at least 95% nucleic acid homology to the PG1 gene is obtained and identified. Similarly, by using progressively less stringent hybridization conditions one can obtain and identify nucleic acids having at least 90%, at least 85%, at least 80% or at least 75% homology to the PG1 gene.

To determine whether a clone encodes a protein having a given amount of homology to the PG1 protein, the amino acid sequence of the PG1 protein is compared to the amino acid sequence encoded by the hybridizing nucleic acid. Homology is determined to exist when an amino acid sequence in the PG1 protein is closely related to an amino acid sequence in the hybridizing nucleic acid. A sequence is closely related when it is identical to that of the PG1 sequence or when it contains one or more amino acid substitutions therein in which amino acids having similar characteristics have been substituted for one another. Using the above methods, one can obtain nucleic acids encoding proteins having at least 95%, at least 90%, at least 85%, at least 80% or at least 75% homology to the proteins encoded by the PG1 probe.

25 <u>Isolation and Use of Mutant or Low Frequency PG1 Alleles from Mammalian Prostate Tumor Tissues</u> and Cell lines

A single mutant PG1 gene was isolated from a human prostate cancer cell line. The nucleic acid sequence and amino acid sequence of this mutant PG1 are disclosed in SEQ IN NOs: 69 and 70, respectively. This mutant was found to contain a stop codon at codon position number 229, and therefore results in a truncated gene product of only 228 amino acids. The present invention encompasses purified or isolated nucleic acids comprising at least 8, 10, 12, 15, 20, or 25 consecutive nucleotides of SEQ ID NO: 69, preferably containing the mutation in codon number 229. A preferred embodiment of the present invention encompasses purified or isolated nucleic acids comprising at least 8, 10, 12, 15, 20, or 25 consecutive nucleotides of SEQ ID NO: 71.

35 The present invention is also directed to methods of determining whether an individual is at risk of developing prostate cancer at a later date or whether said individual suffers from prostate

PCT/IB98/02133

cancer as a result of a mutation in the PG1 gene comprising: obtaining a nucleic acid sample from said individual; and determining whether the nucleotides present at one or more of the polymorphic bases in the sequences selected from the group consisting of SEQ ID NOs: 69 and 71 are indicative of a risk of developing prostate cancer at a later date or indicative of prostate cancer resulting from a mutation in the PG1 gene. The present invention also includes purified or isolated nucleic acids encoding at least 4, 8, 10, 12, 15, or 20 consecutive amino acids of the polypeptide of SEQ ID NO: 70, preferably including the carboxy terminus of said polypeptide. The isolated or purified polypeptides of the invention include polypeptides comprising at least 4, 8, 10, 12, 15, or 20 consecutive amino acids of the polypeptide of SEQ ID NO: 70, preferably including the carboxy terminus of said polypeptide.

V. DIAGNOSIS OF INDIVIDUALS AT RISK FOR DEVELOPING PROSTATE CANCER OR INDIVIDUALS SUFFERING FROM PROSTATE CANCER AS A RESULT OF A MUTATION IN THE PG1 GENE

Individuals may then be screened for the presence of polymorphisms in the PG1 gene or protein which are associated with a detectable phenotype such as cancer, prostate cancer or other conditions as described in Example 13, below. The individuals is screened while they are asymptomatic to determine their risk of developing cancer, prostate cancer or other conditions at a subsequent time. Alternatively, individuals suffering from cancer, prostate cancer or other conditions is screened for the presence of polymorphisms in the PG1 gene or protein in order to determine whether therapies which target the PG1 gene or protein should be applied.

Example 14

Nucleic acid samples are obtained from a symptomatic or asymptomatic individual. The nucleic acid samples is obtained from blood cells as described above or is obtained from other tissues or organs. For individuals suffering from prostate cancer, the nucleic acid sample is obtained from the tumor. The nucleic acid sample may comprise DNA, RNA, or both. The nucleotides at positions in the PG1 gene where mutations lead to prostate cancer or other detectable phenotypes are determined for the nucleic acid sample.

In one embodiment, a PCR amplification is conducted on the nucleic acid sample as described above to amplify regions in which polymorphisms associated with prostate cancer or other detectable phenotypes have been identified. The amplification products are sequenced to determine whether the individual possesses one or more PG1 polymorphisms associated with prostate cancer or other detectable phenotypes.

Alternatively, the nucleic acid sample is subjected to microsequencing reactions as described above to determine whether the individual possesses one or more PG1 polymorphisms associated with prostate cancer or another detectable phenotype resulting from a mutation in the PG1 gene.

In another embodiment, the nucleic acid sample is contacted with one or more allele specific oligonucleotides which specifically hybridize to one or more PG1 alleles associated with prostate cancer or another detectable phenotype. The nucleic acid sample is also contacted with a second PG1 oligonucleotide capable of producing an amplification product when used with the allele specific oligonucleotide in an amplification reaction. The presence of an amplification product in the amplification reaction indicates that the individual possesses one or more PG1 alleles associated with prostate cancer or another detectable phenotype.

Determination of PG1 Expression Levels

As discussed above, PG1 polymorphisms associated with cancer, prostate cancer or other detectable phenotypes may exert their effects by increasing, decreasing, or eliminating PG1 expression, or in altering the frequency of various transcription species. Accordingly, PG1 expression levels in individuals suffering from cancer, prostate cancer or other detectable phenotypes is compared to those of unaffected individuals to determine whether over-expression, under-expression, loss of expression, or changes in the relative frequency of transcription species of PG1 causes cancer, prostate cancer or another detectable phenotype. Individuals is tested to determine whether they are at risk of developing cancer, or prostate cancer at a subsequent time or whether they suffer from prostate cancer resulting from a mutation in the PG1 gene by determining whether they exhibit a level of PG1 expression associated with prostate cancer. Similarly, individuals is tested to determine whether they suffer from another PG1 mediated detectable phenotype or whether they are at risk of suffering from such a condition at a subsequent time.

Expression levels in nucleic acid samples from affected and unaffected individuals is determined by performing Northern blots using detectable probes derived from the PG1 gene or the PG1 cDNA. A variety of conventional Northern blotting procedures is used to detect and quantitate PG1 expression and the frequencies of the various transcription species of PG1, including those disclosed in Current Protocols in Molecular Biology, John Wiley 503 Sons, Inc. 1997 and Sambrook et al. Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, 1989.

Alternatively, PG1 expression levels is determined as described in Example 15, below.

Example 15

described in International Patent Application No. WO 97/05277. Briefly, the PG1 cDNA or the PG1 genomic DNA described above, or fragments thereof, is inserted at a cloning site immediately downstream of a bacteriophage (T3, T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the PG1 insert comprises at least 100 or more consecutive nucleotides of the genomic DNA sequence of SEQ ID NO: 1 or the cDNA sequences of SEQ ID NO: 3. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (i.e. biotin-UTP and

WO 99/32644 PCT/IB98/02133

DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50• C for 16 hours in an 80% formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (i.e. RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

Quantitative analysis of PG1 gene expression may also be performed using arrays as described in Sections II and X,. As used here, the term array means an arrangement of a plurality of nucleic acids 10 of sufficient length to permit specific detection of expression of PG1 mRNAs capable of hybridizing thereto. For example, the arrays may contain a plurality of nucleic acids derived from genes whose expression levels are to be assessed. The arrays may include the PG1 genomic DNA of SEQ ID NO:179, the PG1 cDNA of SEQ ID NO:3 or the sequences complementary thereto or fragments thereof. The array may contain some or all of the known alternative splice or transcription species of PG1, including the species in SEQ ID NOs: 3, and 112-124 to determine the relative frequency of particular transcription species. Alternatively, the array may contain polynucleotides which overlap all of the potential splice junctions, including, for example SEQ ID NOs: 137-178, so that the frequency of particular splice junctions can be determined and correlated with traits or used in diagnostics just as expressions levels are. Preferably, the fragments are at least 15 nucleotides in length. In other embodiments, the fragments are at least 25 nucleotides in length. In some embodiments, the fragments are at least 50 nucleotides in length. More preferably, the fragments are at least 100 nucleotides in length. In another preferred embodiment, the fragments are more than 100 nucleotides in length. In some embodiments the fragments is more than 500 nucleotides in length.

For example, quantitative analysis of PG1 gene expression is performed with a complementary DNA microarray as described by Schena et al. (Science 270:467-470, 1995; Proc. Natl. Acad. Sci. U.S.A. 93:10614-10619, 1996). Full length PG1 cDNAs or fragments thereof are amplified by PCR and arrayed from a 96-well microtiter plate onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95. C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25. C.

Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm² microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60° C. Arrays are washed for 5 min at 25° C in low stringency wash buffer (1 x SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1 x SSC/0.2% SDS). Arrays are scanned in 0.1 x SSC using a fluorescence laser scanning device fitted with

25

30

a custom filter set. Accurate differential expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of PG1 gene expression may also be performed with full length PG1 cDNAs or fragments thereof in complementary DNA arrays as described by Pietu et al. (Genome 5 Research 6:492-503, 1996). The full length PG1 cDNA or fragments thereof is PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by phospho-imaging or autoradiography. Duplicate experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis using the PG1 genomic DNA, the PG1 cDNA, or fragments thereof can be done through high density nucleotide arrays as described by Lockhart et al. (Nature Biotechnology 14: 1675-1680, 1996) and Sosnowsky et al. (Proc. Natl. Acad. Sci. 94:1119-1123, 1997). Oligonucleotides of 15-50 nucleotides from the sequences of the PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, 112-124 or the sequences complementary thereto, are synthesized 15 directly on the chip (Lockhart et al., supra) or synthesized and then addressed to the chip (Sosnowski et al., supra).

PG1 cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin or fluorescent dye, are synthesized from the appropriate mRNA population and then randomly fragmented to an average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After 20 washing as described in Lockhart et al., supra and application of different electric fields (Sosnowsky et al., Proc. Natl. Acad. Sci. 94:1119-1123)., the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed. Comparative analysis of the intensity of the signal originating from cDNA probes on the same target oligonucleotide in different cDNA samples indicates a differential expression of PG1 mRNA.

The above methods may also be used to determine whether an individual exhibits a PG1 expression pattern associated with cancer, prostate cancer or other diseases. In such methods, nucleic acid samples from the individual are assayed for PG1 expression as described above. If a PG1 expression pattern associated with cancer, prostate cancer, or another disease is observed, an appropriate diagnosis is rendered and appropriate therapeutic techniques which target the PG1 gene or protein is applied.

The above methods may also be applied using allele specific probes to determine whether an individual possesses a PG1 allele associated with cancer, prostate cancer, or another disease. In such approaches, one or more allele specific oligonucleotides containing polymorphic nucleotides in the PG1 gene which are associated with prostate cancer are fixed to a microarray. The array is contacted with a nucleic acid sample from the individual being tested under conditions which permit allele specific hybridization of the sample nucleic acid to the allele specific PG1 probes. Hybridization of the sample nucleic acid to one or more of the allele specific PG1 probes indicates that the individual suffers from prostate cancer caused by the PG1 gene or that the individual is at risk for developing prostate cancer at a subsequent time. Alternatively, any of the genotyping methods described in Section X is utilized..

Use of the Biallelic Markers Of The Invention In Diagnostics

The biallelic markers of the present invention can also be used to develop diagnostics tests capable of identifying individuals who express a detectable trait as the result of a specific genotype or individuals whose genotype places them at risk of developing a detectable trait at a subsequent time.

The diagnostic techniques of the present invention may employ a variety of methodologies to determine whether a test subject has a biallelic marker pattern associated with an increased risk of developing a detectable trait or whether the individual suffers from a detectable trait as a result of a particular mutation, including methods which enable the analysis of individual chromosomes for haplotyping, such as family studies, single sperm DNA analysis or somatic hybrids. The trait analyzed using the present diagnostics is any detectable trait, cancer, prostate cancer or another disease, a response to an anti-cancer, or anti-prostate cancer, or side effects to an anti-cancer or anti-prostate cancer agent. Diagnostics, which analyze and predict response to a drug or side effects to a drug, is used to determine whether an individual should be treated with a particular drug. For example, if the diagnostic indicates a likelihood that an individual will respond positively to treatment with a particular drug, the drug is administered to the individual. Conversely, if the diagnostic indicates that an individual is likely to respond negatively to treatment with a particular drug, an alternative course of treatment is prescribed. A negative response is defined as either the absence of an efficacious response or the presence of toxic side effects.

Clinical drug trials represent another application for the markers of the present invention. One or more markers indicative of response to an anti-cancer or anti-prostate cancer agent or to side effects to an anti-cancer or anti-prostate cancer agent is identified using the methods described in Section XI, below. Thereafter, potential participants in clinical trials of such an agent is screened to identify those individuals most likely to respond favorably to the drug and exclude those likely to experience side effects. In that way, the effectiveness of drug treatment is measured in individuals who respond positively to the drug, without lowering the measurement as a result of the inclusion of individuals who are unlikely to respond positively in the study and without risking undesirable safety problems. Preferably, in such diagnostic methods, a nucleic acid sample is obtained from the individual and this sample is genotyped using methods described in Section X.

Another aspect of the present invention relates to a method of determining whether an individual is at risk of developing a trait or whether an individual expresses a trait as a consequence of possessing a particular trait-causing allele. The present invention relates to a method of determining whether an individual is at risk of developing a plurality of traits or whether an individual expresses a plurality of traits as a result of possessing a particular trait-causing allele. These methods involve

20

30

35

obtaining a nucleic acid sample from the individual and determining whether the nucleic acid sample contains one or more alleles of one or more biallelic markers indicative of a risk of developing the trait or indicative that the individual expresses the trait as a result of possessing a particular trait-causing allele.

As described herein, the diagnostics is based on a single biallelic marker or a group of biallelic markers.

VI. ASSAYING THE PG1 PROTEIN FOR INVOLVEMENT IN RECEPTOR/LIGAND INTERACTIONS

The expressed PG1 protein or portion thereof is evaluated for involvement in receptor/ligand interactions as described in Example 16 below.

Example 16

The proteins encoded by the PG1 gene or a portion thereof may also be evaluated for their involvement in receptor/ligand interactions. Numerous assays for such involvement are familiar to those skilled in the art, including the assays disclosed in the following references: Chapter 7.28 (Measurement of Cellular Adhesion under Static Conditions 7.28.1-7.28.22) in Current Protocols in Immunology, J.E. Coligan et al. Eds. Greene Publishing Associates and Wiley-Interscience; Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160, 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995; Gyuris et al., Cell 75:791-803, 1993.

For example, the proteins of the present invention may demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as sclectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The PG1 protein or portions thereof described above is used in drug screening procedures to identify molecules which are agonists, antagonists, or inhibitors of PG1 activity. The PG1 protein or portion thereof used in such analyses is free in solution or linked to a solid support. Alternatively, PG1 protein or portions thereof can be expressed on a cell surface. The cell may naturally express the PG1 protein or portion thereof or, alternatively, the cell may express the PG1 protein or portion thereof from an expression vector such as those described below.

In one method of drug screening, eukaryotic or prokaryotic host cells which are stably transformed with recombinant polynucleotides in order to express the PG1 protein or a portion thereof are used in conventional competitive binding assays or standard direct binding assays. For example, the formation of a complex between the PG1 protein or a portion thereof and the agent being tested is measured in direct binding assays. Alternatively, the ability of a test agent to prevent formation of a complex between the PG1 protein or a portion thereof and a known ligand is measured.

Alternatively, the high throughput screening techniques disclosed in the published PCT application WO 84/03564, is used. In such techniques, large numbers of small peptides to be tested for PG1 binding activity are synthesized on a surface and affixed thereto. The test peptides are contacted with the PG1 protein or a portion thereof, followed by a wash step. The amount of PG1 protein or portion thereof which binds to the test compound is quantitated using conventional techniques.

In some methods, PG1 protein or a portion thereof is fixed to a surface and contacted with a test compound. After a washing step, the amount of test compound which binds to the PG1 protein or portion thereof is measured.

In another approach, the three dimensional structure of the PG1 protein or a portion thereof may be determined and used for rational drug design.

Alternatively, the PG1 protein or a portion thereof is expressed in a host cell using expression vectors such as those described herein. The PG1 protein or portion thereof is an isotype which is associated with prostate cancer or an isotype which is not associated with prostate cancer. The cells 20 expressing the PG1 protein or portion thereof are contacted with a series of test agents and the effects of the test agents on PG1 activity are measured. Test agents which modify PG1 activity is employed in therapeutic treatments.

The above procedures may also be applied to evaluate mutant PG1 proteins responsible for a detectable phenotype.

Identification of Proteins which Interact with the PG1 Protein

Proteins which interact with the PG1 protein is identified as described in Example 17, below.

Example 17

Proteins which interact with the PG1 protein or a portion thereof, is identified using two hybrid systems such as the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described in the manual accompanying the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech), nucleic acids encoding the PG1 protein or a portion thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. cDNAs in a cDNA library which encode proteins which might interact with the polypeptides encoded by the nucleic acids encoding the PG1 protein or a portion thereof are inserted into a second 35 expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which

15

WO 99/32644 PCT/IB98/02133

selects for expression of selectable markers on each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain plasmids encoding proteins which interact with the polypeptide encoded by the nucleic acid inserts.

Alternatively, the system described in Lustig et al., Methods in Enzymology 283: 83-99 (1997), is used for identifying molecules which interact with the PG1 protein or a portion thereof. In such systems, in vitro transcription reactions are performed on vectors containing an insert encoded the PG1 protein or a portion thereof cloned downstream of a promoter which drives in vitro transcription. The resulting mRNA is introduced into Xenopus laevis oocytes. The oocytes are then assayed for a desired activity.

Alternatively, the in vitro transcription products produced as described above is translated in vitro. The in vitro translation products can be assayed for a desired activity or for interaction with a known polypeptide.

The system described in U.S. Patent No. 5,654,150 may also be used to identify molecules which interact with the PG1 protein or a portion thereof. In this system, pools of cDNAs are transcribed and translated in vitro and the reaction products are assayed for interaction with a known polypeptide or antibody.

20 by a variety of additional techniques. In one method, affinity columns containing the PG1 protein or a portion thereof can be constructed. In some versions of this method the affinity column contains chimeric proteins in which the PG1 protein or a portion thereof is fused to glutathione S-transferase. A mixture of cellular proteins or pool of expressed proteins as described above is applied to the affinity column. Proteins interacting with the polypeptide attached to the column can then be isolated and analyzed on 2-D electrophoresis gel as described in Ramunsen et al. Electrophoresis, 18, 588-598 (1997). Alternatively, the proteins retained on the affinity column can be purified by electrophoresis based methods and sequenced. The same method can be used to isolate antibodies, to screen phage display products, or to screen phage display human antibodies.

Proteins interacting with the PG1 protein or portions thereof can also be screened by using an 30 Optical Biosensor as described in Edwards et Leatherbarrow, Analytical Biochemistry, 246, 1-6 (1997). The main advantage of the method is that it allows the determination of the association rate between the protein and other interacting molecules. Thus, it is possible to specifically select interacting molecules with a high or low association rate. Typically a target molecule is linked to the sensor surface (through a carboxymethl dextran matrix) and a sample of test molecules is placed in contact with the target molecules. The binding of a test molecule to the target molecule causes a change in the refractive index and/ or thickness. This change is detected by the Biosensor provided it

occurs in the evanescent field (which extend a few hundred nanometers from the sensor surface). In these screening assays, the target molecule can be the PG1 protein or a portion thereof and the test sample can be a collection of proteins extracted from tissues or cells, a pool of expressed proteins, combinatorial peptide and/ or chemical libraries, or phage displayed peptides. The tissues or cells 5 from which the test proteins are extracted can originate from any species.

In other methods, a target protein is immobilized and the test population is the PG1 protein or a portion thereof.

To study the interaction of the PG1 protein or a portion thereof with drugs, the microdialysis coupled to HPLC method described by Wang et al., Chromatographia, 44, 205-208(1997) or the 10 affinity capillary electrophoresis method described by Busch et al., J. Chromatogr. 777:311-328 (1997).

The above procedures may also be applied to evaluate mutant PG1 proteins responsible for a detectable phenotype.

VII. PRODUCTION OF ANTIBODIES AGAINST PG1 POLYPEPTIDES

Any PG1 polypeptide or whole protein (SEQ ID NOs: 4, 5, 70, 74, 125-136) whether human, mouse or mammalian is used to generate antibodies capable of specifically binding to expressed PG1 protein or fragments thereof as described in Example 16, below. The antibodies is capable of binding the full length PG1 protein. PG1 proteins which result from naturally occurring mutant, particularly functional mutants of PG1, including SEQ ID NO: 70, which may used in the production of antibodies. The present invention also contemplates the use of polypeptides comprising a contiguous stretch of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 50, or 100 amino acids of any PG1 protein in the manufacture of antibodies. In a preferred embodiment the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the PG1 protein sequence. For instance, polypeptides that contain either the Arg and His residues at amino acid position 184, and polypeptides that contain either the Arg or Ile residue at amino acid position 293 of the SEQ ID NO: 4 in said contiguous stretch are particularly preferred embodiments of the invention and useful in the manufacture of antibodies to detect the presence and absence of these mutations. Similarly, polypeptides with a carboxy terminus at position 228 is a particularly preferred 30 embodiment of the invention and useful in the manufacture of antibodies to detect the presence and absence of the mutation shown in SEQ'ID NOs: 69 and 70. Similarly, polypeptides that that contain an peptide sequences of 8, 10, 12, 15, or 25 amino acids encoded over a naturally-occurring splice junction (the point at which two human PG1 exon (SEQ ID NOs: 100-111) are covalently linked) in said contiguous stretch are particularly preferred embodiments and useful in the manufacture of antibodies to detect the presence, localization, and quantity of the various protein products of the PG1 35 alternative splice species.

Alternatively, the antibodies is screened so as to isolate those which are capable of binding an epitope-containing fragment of at least 8, 10, 12, 15, 20, 25, or 30 amino acids of a human, mouse, or mammalian PG1 protein, preferably a sequence selected from SEQ ID NOs: 4, 5, 70, 74, or 125-136.

Antibodies may also be generated which are capable of specifically binding to a given isoform of the PG1 protein. For example, the antibodies is capable of specifically binding to an isoform of the PG1 protein which causes prostate cancer or another detectable phenotype which has been obtained as described above and expressed from an expression vector as described above. Alternatively, the antibodies is capable of binding to an isoform of the PG1 protein which does not cause prostate cancer. Such antibodies is used in diagnostic assays in which protein samples from an individual are evaluated for the presence of an isoform of the PG1 protein which causes cancer or another detectable phenotype using techniques such as Western blotting or ELISA assays.

Non-human animals or mammals, whether wild-type or transgenic, which express a different species of PG1 than the one to which antibody binding is desired, and animals which do not express PG1 (i.e. an PG1 knock out animal as described in Section VIII.) are particularly useful for preparing antibodies. PG1 knock out animals will recognize all or most of the exposed regions of PG1 as foreign antigens, and therefore produce antibodies with a wider array of PG1 epitopes. The humoral immune system of animals which produce a species of PG1 that resembles the antigenic sequence will preferentially recognize the differences between the animal's native PG1 species and the antigen sequence, and produce antibodies to these unique sites in the antigen sequence.

20 Example 18

Substantially pure protein or polypeptide is isolated from transfected or transformed cells containing an expression vector encoding the PG1 protein or a portion thereof as described in Example 11. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes in the PG1 protein or a portion thereof can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., Nature 256:495 (1975) or derivative methods thereof. Also see Harlow, E., and D. Lane. 1988. Antibodies A Laboratory Manual. Cold Spring Harbor Laboratory. pp. 53-242.

Briefly, a mouse is repetitively inoculated with a few micrograms of the PG1 protein or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are

20

identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., Meth. Enzymol. 70:419 (1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al. Basic Methods in Molecular Biology Elsevier, New York. Section 21-2.

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes in the PG1 protein or a portion thereof can be prepared by immunizing suitable non-human animal with the PG1 protein or a portion thereof, which can be unmodified or modified to enhance immunogenicity. A suitable nonhuman animal is preferably a non-human mammal is selected, usually a mouse, rat, rabbit, goat, or horse. Alternatively, a crude preparation which has been enriched for PG1 concentration can be used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant (e.g. aluminum hydroxide, RIBI, etc.) which is known in the art. In addition the protein, fragment or preparation can be pretreated with an agent which will increase antigenicity, such agents are known in the art and include, for example, methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, and keyhole limpet hemocyanin (KLH). Serum from the immunized animal is collected, treated and tested according to known procedures. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies can be purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and 25 Walker (1987). An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. J. Clin. Endocrinol. Metab. 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 • M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980).

Antibody preparations prepared according to either the monoclonal or the polyclonal protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances

in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

VIII. VECTORS AND THE USES OF POLYNUCLEOTIDES IN CELLS, ANIMALS, AND 5 HUMANS

The nucleic acids of the invention include expression vectors, amplification vectors, PCRsuitable polynucleotide primers, and vectors which are suitable for the introduction of a polynucleotide of the invention into an embryonic stem cells for the production of transgenic nonhuman animals. In addition, vectors which are suitable for the introduction of a polynucleotide of the 10 invention into cells, organs and individuals, including human individuals, for the purposes of gene therapy to reduce the severity of or prevent genetic diseases associated with functional mutations in PG1 genes are encompassed by the present invention. Functional mutations in PG1 genes which are suitable as targets for the gene therapy and transgenic vectors and methods of the invention include, but are not limited to, mutations in the coding region of the PG1 gene which affect the amino acid sequence of the PG1 gene's product, mutations in the promoter or other regulatory regions which affect the levels of PG1 expression, mutations in the PG1 splice sites which affect length of the PG1 gene product or the relative frequency of PG1 alternative splicing species, and any other mutation which in any way affects the level or quality of PG1 expression or activity. The gene therapy methods can be achieved by targeting vectors and method for changing a mutant PG1 gene into a 20 wild-type PG1 gene in a embryonic stem cell or somatic cell. Alternatively, the present invention also encompasses methods and vectors for introducing the expression of wild-type PG1 sequences without the disruption of any mutant PG1 which already reside in the cell, organ or individual.

The invention also embodies amplification vectors, which comprise a polynucleotide of the invention, and an origin of replication. Preferably, such amplification vectors further comprise restriction endonuclease sites flanking the polynucleotide, so as to facilitate cleavage and purification of the polynucleotides from the remainder of the amplification vector, and a selectable marker, so as to facilitate amplification of the amplification vector. Most preferably, the restriction endonuclease sites in the amplification vector are situated such that cleavage at those site would result in no other amplification vector fragments of a similar size.

Thus, such an amplification vector is transfected into a host cell compatible with the origin of replication of said amplification vector, wherein the host cell is a prokaryotic or eukaryotic cell, preferably a mammalian, insect, yeast, or bacterial cell, most preferably an Escherichia coli cell. The resulting transfected host cells is grown by culture methods known in the art, preferably under selection compatible with the selectable marker (e.g., antibiotics). The amplification vectors can be isolated and purified by methods known in the art (e.g., standard plasmid prep procedures). The polynucleotide of the invention can be cleaved with restriction enzymes that specifically cleave at the

30

WO 99/32644 PCT/IB98/02133

restriction endonuclease sites flanking the polynucleotide, and the double-stranded polynucleotide fragment purified by techniques known in the art, including gel electrophoresis.

Alternatively linear polynucleotides comprising a polynucleotide of the invention is amplified by PCR. The PCR method is well known in the art and described in, e.g., U.S. Patent Nos. 4,683,195 and 4,683,202 and Saiki, R et al. 1988. Science 239:487-491, and European patent applications 86302298.4, 86302299.2 and 87300203.4, as well as Methods in Enzymology 1987 155:335-350.

The polynucleotides of the invention can also be derivatized in various ways, including those appropriate for facilitating transfection and/or gene therapy. The polynucleotides can be derivatized by attaching a nuclear localization signal to it to improve targeted delivery to the nucleus. One well-characterized nuclear localization signal is the heptapeptide PKKKRKV (pro-lys-lys-arg-lys-val). Preferably, in the case of polynucleotides in the form of a closed circle, the nuclear localization signal is attached via a modified loop nucleotide or spacer that forms a branching structure.

If it is to be used in vivo, the polynucleotide of the invention is derivatized to include ligands and/or delivery vehicles which provide dispersion through the blood, targeting to specific cell types, or permit easier transit of cellular barriers. Thus, the polynucleotides of the invention is linked or combined with any targeting or delivery agent known in the art, including but not limited to, cell penetration enhancers, lipofectin, liposomes, dendrimers, DNA intercalators, and nanoparticles. In particular, nanoparticles for use in the delivery of the polynucleotides of the invention are particles of less than about 50 nanometers diameter, nontoxic, non-antigenic, and comprised of albumin and surfactant, or iron as in the nanoparticle particle technology of SynGenix. In general the delivery vehicles used to target the polynucleotides of the invention may further comprise any cell specific or general targeting agents known in the art, and will have a specific trapping efficiency to the target cells or organs of from about 5 to about 35%.

The polynucleotides of the invention is used ex vivo in a gene therapy method for obtaining cells or organs which produce wild-type PG1 or PG1 proteins which have been selectively mutated. The cells are created by incubation of the target cell with one or more of the above-described polynucleotides under standard conditions for uptake of nucleic acids, including electroporation or lipofection. In practicing an ex vivo method of treating cells or organs, the concentration of polynucleotides of the invention in a solution prepare to treat target cells or organs is from about 0.1 to about 100 µM, preferably 0.5 to 50 µM, most preferably from 1 to 10 µM.

Alternatively, the oligonucleotides can be modified or co-administered for targeted delivery to the nucleus. Improved oligonucleotide stability is expected in the nucleus due to: (1) lower levels of DNases and RNases; and (2) higher oligonucleotide concentrations due to lower total volume.

Alternatively, the polynucleotides of the invention can be covalently bonded to biotin to form a biotin-polynucleotide prodrug by methods known in the art, and co-administered with a receptor ligand bound to avidin or receptor specific antibody bound to avidin, wherein the receptor is capable

of causing uptake of the resulting polynucleotide-biotin-avidin complex into the cells. Receptors that cause uptake are known to those of skill in the art.

The invention encompasses vectors which are suitable for the introduction of a polynucleotide of the invention into an embryonic stem cell for the production of transgenic non-human animals, which in turn result in the expression of recombinant PG1 in the transgenic animal. Any appropriate vector system can be used for the introduction and expression of PG1 in transgenic animals, including for example yeast artificial chromosomes (YAC), bacterial artificial chromosomes (BAC), bacteriophage P1, and other vectors known in the art which are able to accommodate sufficiently large inserts to encode the PG1 protein or desired fragments thereof. Selected alterations, additions and deletions in the PG1 gene may optionally be achieved by site-directed mutagenesis. Once an appropriate vector system is chosen, the site-directed mutagenesis process may then be conducted by techniques well known in the art, and the fragment be returned and ligated to the larger vector from which it was cleaved. For site directed mutagenesis methods see, for example, Kunkel, T. 1985. Proc. Natl. Acad. Sci. U.S.A. 82:488; Bandeyar, M. et al. 1988. Gene 65: 129-133; Nelson, M., and M. McClelland 1992. Methods Enzymol. 216:279-303; Weiner, M. 1994. Gene 151: 119-123; Costa, G. and M. Weiner. 1994. Nucleic Acids Res. 22: 2423; Hu, G. 1993. DNA and Cell Biology 12:763-770; and Deng, W. and J. Nickoff. 1992. Anal. Biochem. 200:81.

Briefly, the transgenic technology used herein involves the inactivation, addition or replacement of a portion of the PG1 gene or the entire gene. For example the present technology 20 includes the addition of PG1 genes with or without the inactivation of the non-human animal's native PG1 genes, as described in the preceding two paragraphs and in the Examples. The invention also encompasses the use of vectors, and the vectors themselves which target and modify an existing human PG1 gene in a stem cell, whether it is contained in a non-human animal cell where it was previously introduced into the germ line by transgenic technology or it is a native PG1 gene in a human pluripotent or somatic cell. This transgene technology usually relies on homologous recombination in a pluripotent cell that is capable of differentiating into germ cell tissue. A DNA construct that encodes an altered region of the non-human animal's PG1 gene that contains, for instance a stop codon to destroy expression, is introduced into the nuclei of embryonic stem cells. Preferably mice are used for this transgenic work. In a portion of the cells, the introduced DNA recombines with the endogenous copy of the cell's gene, replacing it with the altered copy. Cells containing the newly engineered genetic alteration are injected in a host embryo of the same species as the stem cell, and the embryo is reimplanted into a recipient female. Some of these embryos develop into chimeric individuals that posses germ cells entirely derived from the mutant cell line. Therefore, by breeding the chimeric progeny it is possible to obtain a new strain containing the 35 introduced genetic alteration. See Capecchi 1989. Science. 244:1288-1292 for a review of this procedure.

The present invention encompasses the polynucleotides described herein, as well as the methods for making these polynucleotides including the method for creating a mutation in a human PG1 gene. In addition, the present invention encompasses cells which comprise the polynucleotides of the invention, including but not limited to amplification host cells comprising amplification vectors 5 of the invention. Furthermore the present invention comprises the embryonic stem cells and transgenic non-human animals and mammals described herein which comprise a gene encoding a human PG1 protein.

DNA construct that enables directing temporal and spatial gene expression in recombinant host cells and in transgenic animals

In order to study the physiological and phenotype consequences of a lack of synthesis of the PG1 protein, both at the cellular level and at the multi-cellular organism level, in particular as regards to disorders related to abnormal cell proliferation, notably cancers, the invention also encompasses DNA constructs and recombinant vectors enabling a conditional expression of a specific allele of the PG1 genomic sequence or cDNA and also of a copy of this genomic sequence or cDNA harboring 15 substitutions, deletions, or additions of one or more bases as regards to the PG1 nucleotide sequence of SEQ ID NOs: 3, 112-125, 179, 182-184, or a fragment thereof, these base substitutions, deletions or additions being located either in an exon, an intron or a regulatory sequence, but preferably in a 5'regulatory sequence of a mammalian PG1 gene, more preferably SEQ ID NO: 180 or in an exon of the PG1 genomic sequence or within the PG1 cDNA of SEQ ID NOs 3, 112-125, or 184.

A first preferred DNA construct is based on the tetracycline resistance operon tet from E. coli transposon Tn110 for controlling the PG1 gene expression, such as described by Gossen M. et al., 1992, Proc. Natl. Acad. Sci. USA, 89: 5547-5551; Gossen M. et al., 1995, Science, 268: 1766-1769; and Furth P.A. et al., 1994, Proc. Natl Acad. Sci USA, 91: 9302-9306. Such a DNA construct contains seven tet operator sequences from Tn10 (tetop) that are fused to either a minimal promoter or a 5'-regulatory sequence of the PG1 gene, said minimal promoter or said PG1 regulatory sequence being operably linked to a polynucleotide of interest that codes either for a sense or an antisense oligonucleotide or for a polypeptide, including a PG1 polypeptide or a peptide fragment thereof. This DNA construct is functional as a conditional expression system for the nucleotide sequence of interest when the same cell also comprises a nucleotide sequence coding for either the wild type (tTA) or the 30 mutant (rTA) repressor fused to the activating domain of viral protein VP16 of herpes simplex virus, placed under the control of a promoter, such as the HCMVIE1 enhancer/promoter or the MMTV-LTR. Indeed, a preferred DNA construct of the invention will comprise both the polynucleotide containing the tet operator sequences and the polynucleotide containing a sequence coding for the tTA or the rTA repressor.

10

In the specific embodiment wherein the conditional expression DNA construct contains the sequence encoding the mutant tetracycline repressor rTA, the expression of the polynucleotide of interest is silent in the absence of tetracycline and induced in its presence.

DNA constructs allowing homologous recombination: replacement vectors

A second preferred DNA construct will comprise, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised of a PG1 sequence preferably a PG1 genomic sequence; (b) a nucleotide sequence comprising a positive selection marker, such as the marker for neomycin resistance (neo); and (c) a second nucleotide sequence that comprised of a PG1 sequence preferably a PG1 genomic sequence, and is located on the genome downstream the first PG1 nucleotide sequence (a).

In a preferred embodiment, this DNA construct also comprises a negative selection marker located upstream the nucleotide sequence (a) or downstream the nucleotide sequence (b). Preferably, the negative selection marker consists of the thymidine kinase (tk) gene (Thomas K.R. et al., 1986, Cell, 44: 419-428), the hygromycin beta gene (Te Riele et al., 1990, Nature, 348: 649-651), the hprt gene (Van der Lugt et al., 1991, Gene, 105: 263-267; and Reid L.H. et al., 1990, Proc. Natl. Acad. Sci. 15 USA, 87: 4299-4303) or the Diphteria toxin A fragment (Dt-A) gene (Nada S. et al., 1993, Cell, 73: 1125-1135; Yagi T. et al., 1990, Proc. Natl. Acad. Sci. USA, 87: 9918-9922). Preferably, the positive selection marker is located within a PG1 exon sequence so as to interrupt the sequence encoding a PG1 protein.

These replacement vectors are described for example by Thomas K.R. et al., 1986, Cell, 44: 20 419-428; Thomas K.R. et al., 1987, Cell, 51: 503-512; Mansour S.L. et al., 1988, Nature, 336: 348-352; and Koller et al., 1992, Annu. Rev. Immunol., 10: 705-30.

The first and second nucleotide sequences (a) and (c) is located at any point within a PG1 regulatory sequence, an intronic sequence, an exon sequence or a sequence containing both regulatory and/or intronic and/or exon sequences. The length of nucleotide sequences (a) and (c) is determined 25 empirically by one of ordinary skill in the art. Nucleotide sequences (a) and (c) or any length are specifically contemplated in the present invention, however, lengths ranging from 1 kb to 50 kb, preferably from 1 kb to 10 kb, more preferably from 2 kb to 6 kb and most preferably from 2 kb to 4 kb are normally used.

DNA constructs allowing homologous recombination: Cre-loxP system.

These new DNA constructs make use of the site-specific recombination system of the P1 30 phage. The P1 phage possesses a recombinase called Cre which interacts specifically with a 34 base pairs loxP site. The loxP site is composed of two palindromic sequences of 13 bp separated by a 8 bp conserved sequence (Hoess et al., 1986, Nucleic Acids Res., 14: 2287-2300). The recombination by the Cre enzyme between two loxP sites having an identical orientation leads to the deletion of the 35 DNA fragment.

5

The Cre-loxP system used in combination with a homologous recombination technique has been first described by Gu H. et al., 1993, Cell, 73: 1155-1164; and Gu H. et al., 1994, Science, 265: 103-106. Briefly, a nucleotide sequence of interest to be inserted in a targeted location of the genome harbors at least two loxP sites in the same orientation and located at the respective ends of a 5 nucleotide sequence to be excised from the recombinant genome. The excision event requires the presence of the recombinase (Cre) enzyme within the nucleus of the recombinant host cell. The recombinase enzyme is brought at the desired time either by (a) incubating the recombinant host cells in a culture medium containing this enzyme, by injecting the Cre enzyme directly into the desired cell, such as described by Araki K. et al., 1995, Proc. Natl. Acad. Sci. USA, 92: 160-164; or by lipofection of the enzyme into the cells, such as described by Baubonis et al., 1993, Nucleic Acids Res., 21: 2025-2029; (b) transfecting the cell host with a vector comprising the Cre coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter being optionally inducible, said vector being introduced in the recombinant cell host, such as described by Gu H. et al., 1993, Cell, 73: 1155-1164; and Sauer B. et al., 1988, Proc. Natl. Acad. Sci. USA, 85: 5166-5170; (c) introducing in 15 the genome of the host cell a polynucleotide comprising the Cre coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter is optionally inducible, and said polynucleotide being inserted in the genome of the cell host either by a random insertion event or an homologous recombination event, such as described by Gu H. et al., 1994, Science, 265: 103-106.

In the specific embodiment wherein the vector containing the sequence to be inserted in the PG1 gene by homologous recombination is constructed in such a way that selectable markers are flanked by loxP sites of the same orientation, it is possible, by treatment by the Cre enzyme, to eliminate the selectable markers while leaving the PG1 sequences of interest that have been inserted by an homologous recombination event. Again, two selectable markers are needed: a positive selection marker to select for the recombination event and a negative selection marker to select for the homologous recombination event. Vectors and methods using the Cre-loxP system are described by Zou Y.R. et al., 1994, Curr. Biol., 4: 1099-1103.

Thus, a third preferred DNA construct of the invention comprises, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised of a PG1 sequence, preferably a PG1 genomic sequence; (b) a nucleotide sequence comprising a polynucleotide encoding a positive selection marker, such as the marker for neomycin resistance (neo), said nucleotide sequence comprising additionally two sequences defining a site recognized by a recombinase, such as a loxP site, the two sites being placed in the same orientation; and (c) a second nucleotide sequence that is comprised of a PG1 sequence, preferably a PG1 genomic sequence, and is located on the genome downstream of the first PG1 nucleotide sequence (a).

35 The sequences defining a site recognized by a recombinase, such as a loxP site, are preferably located within the nucleotide sequence (b) at suitable locations bordering the nucleotide sequence for

which the conditional excision is sought. In one specific embodiment, two loxP sites are located at each side of the positive selection marker sequence, in order to allow its excision at a desired time after the occurrence of the homologous recombination event.

In a preferred embodiment of a method using the third DNA construct described above, the excision of the polynucleotide fragment bordered by the two sites recognized by a recombinase, preferably two loxP sites, is performed at a desired time, due to the presence within the genome of the recombinant host cell of a sequence encoding the Cre enzyme operably linked to a promoter sequence, preferably an inducible promoter, more preferably a tissue-specific promoter sequence and most preferably a promoter sequence which is both inducible and tissue-specific, such as described by Gu H. et al., 1994, Science, 265: 103-106.

The presence of the Cre enzyme within the genome of the recombinant cell host may result of the breeding of two transgenic animals, the first transgenic animal bearing the PG1-derived sequence of interest containing the loxP sites as described above and the second transgenic animal bearing the Cre coding sequence operably linked to a suitable promoter sequence, such as described by Gu H. et al., 1994, Science, 265: 103-106. Spatio-temporal control of the Cre enzyme expression may also be achieved with an adenovirus based vector that contains the Cre gene thus allowing infection of cells, or in vivo infection of organs, for delivery of the Cre enzyme, such as described by Anton M. et al., 1995, J. Virol., 69: 4600-4606; and Kanegae Y. et al., 1995, Nucl. Acids Res., 23: 3816-3821.

The DNA constructs described above is used to introduce a desired nucleotide sequence of the invention, preferably a PG1 genomic sequence or a PG1 cDNA sequence, and most preferably an altered copy of a PG1 genomic or cDNA sequence, within a predetermined location of the targeted genome, leading either to the generation of an altered copy of a targeted gene (knock-out homologous recombination) or to the replacement of a copy of the targeted gene by another copy sufficiently homologous to allow an homologous recombination event to occur (knock-in homologous recombination).

Nuclear antisense DNA constructs

Preferably, the antisense polynucleotides of the invention have a 3' polyadenylation signal that has been replaced with a self-cleaving ribozyme sequence, such that RNA polymerase II transcripts are produced without poly(A) at their 3' ends, these antisense polynucleotides being incapable of export from the nucleus, such as described by Liu Z. et al., 1994, Proc. Natl. Acad. Sci. USA, 91: 4528-4262. In a preferred embodiment, these PG1 antisense polynucleotides also comprise, within the ribozyme cassette, a histone stem-loop structure to stabilize cleaved transcripts against 3'-5' exonucleolytic degradation, such as described by Eckner R. et al., 1991, EMBO J., 10: 3513-3522.

Expression Vectors

35 The polynucleotides of the invention also include expression vectors. Expression vector systems, control sequences and compatible host are known in the art. For a review of these systems

see, for example, U.S. Patent No. 5,350,671, columns 45-48. Any of the standard methods known to those skilled in the art for the insertion of DNA fragments into a vector is used to construct expression vectors containing a chimeric gene consisting of appropriate transcriptional/translational control signals and the protein coding sequences. These methods may include in vitro recombinant DNA and synthetic techniques and in vivo recombinants (genetic recombination).

Expression of a polypeptide, peptide or derivative, or analogs thereof encoded by a polynucleotide sequence in SEQ ID NOs: 3, 69, 100-112, or 179-184 is regulated by a second nucleic acid sequence so that the protein or peptide is expressed in a host transformed with the recombinant For example, expression of a protein or peptide is controlled by any DNA molecule. promoter/enhancer element known in the art. Promoters which is used to control expression include, but are not limited to, the CMV promoter, the SV40 early promoter region (Bernoist and Chambon, 1981, Nature 290:304-310), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto, et al., 1980, Cell 22:787-797), the herpes thymidine kinase promoter (Wagner et al., 1981, Proc. Natl. Acad. Sci. U.S.A. 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster et al., 1982, Nature 296:39-42); prokaryotic expression vectors such as the beta-lactamase promoter (Villa-Kamaroff, et al., 1978, Proc. Natl. Acad. Sci. U.S.A. 75:3727-3731), or the tac promoter (DeBoer, et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:21-25); see also "Useful proteins from recombinant bacteria" in Scientific American, 1980, 242:74-94; plant expression vectors comprising the nopaline synthetase promoter region (Herrera-Estrella et al., 1983, 20 Nature 303:209-213) or the cauliflower mosaic virus 35S RNA promoter (Gardner, et al., 1981, Nucl. Acids Res. 9:2871), and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase (Herrera-Estrella et al., 1984, Nature 310:115-120); promoter elements from yeast or other fungi such as the Gal 4 promoter, the ADC (alcohol dehydrogenase) promoter, PGK (phosphoglycerol kinase) promoter, alkaline phosphatase promoter, and the following animal transcriptional control regions, which exhibit tissue specificity and have been utilized in transgenic animals: elastase I gene control region which is active in pancreatic acinar cells (Swift et al., 1984, Cell 38:639-646; Ornitz et al., 1986, Cold Spring Harbor Symp. Quant. Biol. 50:399-409; MacDonald, 1987, Hepatology 7:425-515); insulin gene control region which is active in pancreatic beta cells (Hanahan, 1985, Nature 315:115-122), immunoglobulin gene control region which is active in lymphoid cells (Grosschedl et al., 1984, Cell 38:647-658; Adames et al., 1985, Nature 318:533-538; Alexander et al., 1987, Mol. Cell. Biol. 7:1436-1444), mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells (Leder et al., 1986, Cell 45:485-495), albumin gene control region which is active in liver (Pinkert et al., 1987, Genes and Devel. 1:268-276), alpha-fetoprotein gene control region which is active in liver (Krumlauf et al., 1985, Mol. Cell. Biol. 5:1639-1648; Hammer 35 et al., 1987, Science 235:53-58; alpha 1-antitrypsin gene control region which is active in the liver (Kelsey et al., 1987, Genes and Devel. 1:161-171), beta-globin gene control region which is active in myeloid cells (Mogram et al., 1985, Nature 315:338-340; Kollias et al., 1986, Cell 46:89-94; myelin basic protein gene control region which is active in oligodendrocyte cells in the brain (Readhead et al., 1987, Cell 48:703-712); myosin light chain-2 gene control region which is active in skeletal muscle (Sani, 1985, Nature 314:283-286), and gonadotropic releasing hormone gene control region which is active in the hypothalamus (Mason et al., 1986, Science 234:1372-1378).

Other suitable vectors, particularly for the expression of genes in mammalian cells, is selected from the group of vectors consisting of P1 bacteriophages, and bacterial artificial chromosomes (BACs). These types of vectors may contain large inserts ranging from about 80-90 kb (P1 bacteriophage) to about 300 kb (BACs).

P1 bacteriophage

The construction of P1 bacteriophage vectors such as p158 or p158/neo8 are notably described by Sternberg N.L., 1992, Trends Genet., 8: 1-16; and Sternberg N.L., 1994, Mamm. Genome, 5: 397-404. Recombinant P1 clones comprising PG1 nucleotide sequences is designed for inserting large polynucleotides of more than 40 kb (Linton M.F. et al., 1993, J. Clin. Invest., 92: 3029-3037). To generate P1 DNA for transgenic experiments, a preferred protocol is the protocol described by McCormick et al., 1994, Genet. Anal. Tech. Appl., 11: 158-164. Briefly, E. coli (preferably strain NS3529) harboring the P1 plasmid are grown overnight in a suitable broth medium containing 25 µg/ml of kanamycin. The P1 DNA is prepared from the E. coli by alkaline lysis using the Qiagen Plasmid Maxi kit (Qiagen, Chatsworth, CA, USA), according to the manufacturer's instructions. The P1 DNA is purified from the bacterial lysate on two Qiagen-tip 500 columns, using the washing and elution buffers contained in the kit. A phenol/chloroform extraction is then performed before precipitating the DNA with 70% ethanol. After solubilizing the DNA in TE (10 mM Tris-HCl, pH 7.4, 1 mM EDTA), the concentration of the DNA is assessed by spectrophotometry.

When the goal is to express a P1 clone comprising PG1 nucleotide sequences in a transgenic animal, typically in transgenic mice, it is desirable to remove vector sequences from the P1 DNA fragment, for example by cleaving the P1 DNA at rare-cutting sites within the P1 polylinker (SfiI, Notl or Sall). The P1 insert is then purified from vector sequences on a pulsed-field agarose gel, using methods similar using methods similar to those originally reported for the isolation of DNA from YACs (Schedl A. et al., 1993, Nature, 362: 258-261; and Peterson et al., 1993, Proc. Natl. Acad. Sci. USA, 90: 7593-7597). At this stage, the resulting purified insert DNA can be concentrated, if necessary, on a Millipore Ultrafree-MC Filter Unit (Millipore, Bedford, MA, USA – 30,000 molecular weight limit) and then dialyzed against microinjection buffer (10 mM Tris-HCl, pH 7.4; 250 µM EDTA) containing 100 mM NaCl, 30 µM spermine, 70 µM spermidine on a microdyalisis membrane (type VS, 0.025 µM from Millipore). The intactness of the purified P1 DNA insert is assessed by electrophoresis on 1% agarose (Sea Kem GTG; FMC Bio-products) pulse-field gel and staining with ethidium bromide.

PCT/IB98/02133

Bacterial Artificial Chromosomes (BACs)

The bacterial artificial chromosome (BAC) cloning system (Shizuya et al., 1992, Proc. Natl. Acad. Sci. USA, 89: 8794-8797) has been developed to stably maintain large fragments of genomic DNA (100-300 kb) in E. coli. A preferred BAC vector consists of pBeloBAC11 vector that has been 5 described Kim U. J., et al., 1996, Genomics, 34: 213-218. BAC libraries are prepared with this vector using size-selected genomic DNA that has been partially digested using enzymes that permit ligation into either the Bam HI or Hind III sites in the vector. Flanking these cloning sites are T7 and SP6 RNA polymerase transcription initiation sites that can be used to generate end probes by either RNA transcription or PCR methods. After the construction of a BAC library in E. coli, BAC DNA is purified from the host cell as a supercoiled circle. Converting these circular molecules into a linear form precedes both size determination and introduction of the BACs into recipient cells. The cloning site is flanked by two Not I sites, permitting cloned segments to be excised from the vector by Not I digestion. Alternatively, the DNA insert contained in the pBeloBAC11 vector is linearized by treatment of the BAC vector with the commercially available enzyme lambda terminase that leads to the cleavage at the unique cosN site, but this cleavage method results in a full length BAC clone 15 containing both the insert DNA and the BAC sequences.

Host Cells

The PG1 gene expression in human cells is rendered defective, or alternatively it is proceeded with the insertion of a PG1 genomic or cDNA sequence with the replacement of the PG1 gene 20 counterpart in the genome of an animal cell by a PG1 polynucleotide according to the invention. These genetic alterations is generated by homologous recombination events using specific DNA constructs that have been previously described.

One kind of host cell that is used are mammal zygotes, such as murine zygotes. For example, murine zygotes may undergo microinjection with a purified DNA molecule of interest, for example a purified DNA molecule that has previously been adjusted to a concentration range from 1 ng/ml -for BAC inserts- 3 ng/µl -for P1 bacteriophage inserts- in 10 mM Tris-HCl, pH 7.4, 250 µM EDTA containing 100 mM NaCl, 30 μM spermine, and 70 μM spermidine. When the DNA to be microinjected has a large size, polyamines and high salt concentrations can be used in order to avoid mechanical breakage of this DNA, as described by Schedl et al., 1993, Nucleic Acids Res., 21: 4783-4787.

Anyone of the polynucleotides of the invention, including the DNA constructs described herein, is introduced in an embryonic stem (ES) cell line, preferably a mouse ES cell line. ES cell lines are derived from pluripotent, uncommitted cells of the inner cell mass of pre-implantation blastocysts. Preferred ES cell lines are the following: ES-E14TG2a (ATCC No. CRL-1821), ES-D3 (ATCC No. CRL1934 and No. CRL-11632), YS001 (ATCC No. CRL-11776), 36.5 (ATCC No. CRL-11116). To maintain ES cells in an uncommitted state, they are cultured in the presence of growth

30

WO 99/32644 PCT/IB98/02133 95

inhibited feeder cells which provide the appropriate signals to preserve this embryonic phenotype and serve as a matrix for ES cell adherence. Preferred feeder cells consist of primary embryonic fibroblasts that are established from tissue of day 13- day 14 embryos of virtually any mouse strain, that are maintained in culture, such as described by Abbondanzo SJ et al., 1993, Methods in Enzymology, Academic Press, New York, pp. 803-823; and are inhibited in growth by irradiation, such as described by Robertson E., 1987, Embryo-derived stem cell lines. E.J. Robertson Ed. Teratocarcinomas and embrionic stem cells: a practical approach. IRL Press, Oxford, pp. 71, or by the presence of an inhibitory concentration of LIF, such as described by Pease S. and William R.S., 1990, Exp. Cell. Res., 190: 209-211.

10 Transgenic Animals

The terms "transgenic animals" or "host animals" are used herein designate non-human animals that have their genome genetically and artificially manipulated so as to include one of the nucleic acids according to the invention. Preferred animals are non-human mammals and include those belonging to a genus selected from Mus (e.g. mice), Rattus (e.g. rats) and Oryctogalus (e.g. rabbits) which have their genome artificially and genetically altered by the insertion of a nucleic acid according to the invention.

The transgenic animals of the invention all include within a plurality of their cells a cloned recombinant or synthetic DNA sequence, more specifically one of the purified or isolated nucleic acids comprising a PG1 coding sequence, a PG1 regulatory polynucleotide or a DNA sequence encoding an antisense polynucleotide such as described in the present specification.

Preferred transgenic animals according to the invention contains in their somatic cells and/or in their germ line cells a polynucleotide selected from the following group of polynucleotides:

- a) non-native, purified or isolated nucleic acid encoding a PG1 polypeptide, or a polypeptide fragment or variant thereof.
- b) a non-native, purified or isolated nucleic comprising at least 8 consecutive nucleotides of the nucleotide sequence SEQ ID NOs: 179, 182, or 183, a nucleotide sequence complementary; in some embodiments, the length of the fragments can range from at least 8, 10, 15, 20 or 30 to 200 nucleotides, preferably from at least 10 to 50 nucleotides, more preferably from at least 40 to 50 nucleotides of SEQ ID NOs: 179, 182, or 183, or the sequence complementary thereto. In some embodiments, the fragments may comprise more than 200 nucleotides of SEQ ID NOs: 179, 182, or 183, or the sequence complementary thereto.
- c) a non-native, purified or isolated nucleic acid comprising at least 8 consecutive nucleotides of the nucleotide sequence SEQ ID NOs: 3, 69, 112-125 or 184, a sequence complementary thereto or a variant thereof; In some embodiments, the length of the fragments can range from at least 8, 10, 15, 20 or 30 to 200 nucleotides, preferably from at least 10 to 50 nucleotides, more preferably from at least 40 to 50 nucleotides of SEQ ID NOs: 3, 69, 112-125 or 184, or the sequence complementary

thereto. In some embodiments, the fragments may comprise more than 200 nucleotides of SEQ ID NOs: 3, 69, 112-125 or 184, or the sequence complementary thereto.

- d) a non-native, purified or isolated nucleic acid comprising a nucleotide sequence selected from the group of SEQ ID NOs: 100 to 111, a sequence complementary thereto or a fragment or a variant thereof.
 - e) a non-native, purified or isolated nucleic acid comprising a combination of at least two polynucleotides selected from the group consisting of SEQ ID NOs: 100 to 111, or the sequences complementary thereto wherein the polynucleotides are arranged within the nucleic acid, from the 5' end to the 3'end of said nucleic acid, in the same order than in SEQ NOs: 179, 182, or 183.
- f) a non-native, purified or isolated nucleic acid comprising the nucleotide sequence SEQ ID NO: 180, or the sequences complementary thereto or a biologically active fragment or variant of the nucleotide sequence of SEQ ID NO: 180, or the sequence complementary thereto.
- g) a non-native, purified or isolated nucleic acid comprising the nucleotide sequence SEQ ID NO: 181, or the sequence complementary thereto or a biologically active fragment or variant of the nucleotide sequence of SEQ ID NO: 181 or the sequence complementary thereto.
 - h) a polynucleotide consisting of:
 - (1) a nucleic acid comprising a regulatory polynucleotide of SEQ ID NO: 180 or the sequences complementary thereto or a biologically active fragment or variant thereof
 - (2) a polynucleotide encoding a desired polypeptide or nucleic acid.
- 20 (3) Optionally, a nucleic acid comprising a regulatory polynucleotide of SEQ NO: 181, or the sequence complementary thereto or a biologically active fragment or variant thereof.
 - i) a DNA construct as described previously in the present specification.

The transgenic animals of the invention thus contain specific sequences of exogenous genetic material or "non-native" such as the nucleotide sequences described above in detail.

In a first preferred embodiment, these transgenic animals is good experimental models in order to study the diverse pathologies related to cell differentiation, in particular concerning the transgenic animals within the genome of which has been inserted one or several copies of a polynucleotide encoding a native PG1 protein, or alternatively a mutant PG1 protein.

In a second preferred embodiment, these transgenic animals may express a desired polypeptide of interest under the control of the regulatory polynucleotides of the PG1 gene, leading to good yields in the synthesis of this protein of interest, and eventually a tissue specific expression of this protein of interest.

The design of the transgenic animals of the invention is made according to the conventional techniques well known from the one skilled in the art. For more details regarding the production of transgenic animals, and specifically transgenic mice, it is referred to Sandou et al. (1994) and also to US Patents Nos 4,873,191, issued Oct.10, 1989, 5,464,764 issued Nov 7, 1995 and 5,789,215, issued

PCT/IB98/02133

97

Aug 4, 1998.

Transgenic animals of the present invention are produced by the application of procedures which result in an animal with a genome that has incorporated exogenous genetic material. The procedure involves obtaining the genetic material, or a portion thereof, which encodes either a PG1 coding sequence, a PG1 regulatory polynucleotide or a DNA sequence encoding a PG1 antisense polynucleotide such as described in the present specification.

A recombinant polynucleotide of the invention is inserted into an embryonic or ES stem cell line. The insertion is preferably made using electroporation, such as described by Thomas K.R. et al., 1987, Cell, 51: 503-512. The cells subjected to electroporation are screened (e.g. by selection via selectable markers, by PCR or by Southern blot analysis) to find positive cells which have integrated the exogenous recombinant polynucleotide into their genome, preferably via an homologous recombination event. An illustrative positive-negative selection procedure that is used according to the invention is described by Mansour S.L. et al., 1988, Nature, 336: 348-352.

Then, the positive cells are isolated, cloned and injected into 3.5 days old blastocysts from mice, such as described by Bradley A., 1987, Production and analysis of chimaeric mice. In: E.J. Robertson (Ed.), Teratocarcinomas and embryonic stem cells: A practical approach. IRL Press, Oxford, pp.113. The blastocysts are then inserted into a female host animal and allowed to grow to term.

Alternatively, the positive ES cells are brought into contact with embryos at the 2.5 days old 8-16 cell stage (morulae) such as described by Wood S.A. et al., 1993, Proc. Natl. Acad. Sci. USA, 90: 4582-4585; or by Nagy A. et al., 1993, Proc. Natl. Acad. Sci. USA, 90: 8424-8428. The ES cells being internalized to colonize extensively the blastocyst including the cells which will give rise to the germ line. The offspring of the female host are tested to determine which animals are transgenic e.g. include the inserted exogenous DNA sequence and which are wild-type.

Thus, the present invention also concerns a transgenic animal containing a nucleic acid, a recombinant expression vector or a recombinant host cell according to the invention.

Recombinant cell lines derived from the transgenic animals of the invention.

A further object of the invention consists of recombinant host cells obtained from a transgenic animal described herein.

Recombinant cell lines is established in vitro from cells obtained from any tissue of a transgenic animal according to the invention, for example by transfection of primary cell cultures with vectors expressing onc-genes such as SV40 large T antigen, as described by Chou J.Y., 1989, Mol. Endocrinol., 3: 1511-1514; and Shay J.W. et al., 1991, Biochem. Biophys. Acta, 1072: 1-7.

Functional Analysis of the PG1 Poplypeptides In Transgenic Animals

Using different BACs that contain the PG1 gene, we performed FISH experiment on the adenocarcinoma prostatic cell line PC3. Only one signal could be detected showing that this region of chromosome 8 is hemizygous in this tumoral cell line.

To study the function of PG1, it is inactivate by homologous recombination in the remaining allele of PG1 in the PC3 cell line. To inactivate the remaining PG1 allele, a knock-out targeting vector is generated by inserting two genomic DNA fragments of 3.0 and 4.3 kb (that correspond to a sequence upstream of the PG1 promoter and to part of intron 1, respectively) in the pKO Scrambler Neo TK vector (Lexicon ref V1901). Since the targeting vector contains the neomycine resistance 10 gene as well as the Tk gene, homologous recombination is selected by adding geneticin and FIAU to the medium. The promoter, the transcriptional start site, and the first ATG contained in exon 1 on the recombinant allele is deleted by homologous recombination between the targeting vector and the remaining PG1 allele. Accordingly, no coding transcripts is initiated from the recombinant allele. The parental PC3 cells as well as cells hemizygous for the null allele are assessed for their phenotype, their growth rate in liquid culture, their ability to grow in agar (anchorage-independent growth) as well as their ability to form tumors and metastasis when injected subcutaneously in nude mice.

To determine the function of PG1 in the animal, and to generate an animal model for prostate tumorigenesis, mice in which tissue specific inactivation of the PG1 alleles can be induced are generated. For this purpose, the Cre-loxP system is utilized as described above to allow chromosome engineering to be perform directly in the animal.

First, to generate mice with a conditional null allele, two loxP sites are introduced in the murine genome, the first one 5' to the PG1 promoter and the second one 3' to the PG1 exon 1. Alternatively, to generate subtle mutations or to specifically mutate some isoforms, the loxP sites are introduced so that they flank any of the given exons or any potential set of exons. It is important to 25 note that a functional PG1 messenger can be transcribed from these alleles until a recombination is triggered between the loxP sites by the Cre enzyme.

Second, to generate the inducer mice, the Cre gene is introduced in the mouse genome under the control of a tissue specific promoter, for example under the control of the PSA (prostate specific antigen) promoter.

Finally, tissue specific inactivation of the PG1 gene are induced by generating mice 30 containing the Cre transgene that are homozygous for the recombinant PG1 allele.

Gene Therapy

The present invention also comprises the use of the PG1 genomic DNA sequence of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, or nucleic acid encoding a mutant PG1 protein responsible 35 for a detectable phenotype in gene therapy strategies, including antisense and triple helix strategies as described in Examples 19 and 20, below. In antisense approaches, nucleic acid sequences complementary to an mRNA are hybridized to the mRNA intracellularly, thereby blocking the expression of the protein encoded by the mRNA. The antisense sequences may prevent gene expression through a variety of mechanisms. For example, the antisense sequences may inhibit the ability of ribosomes to translate the mRNA. Alternatively, the antisense sequences may block transport of the mRNA from the nucleus to the cytoplasm, thereby limiting the amount of mRNA available for translation. Another mechanism through which antisense sequences may inhibit gene expression is by interfering with mRNA splicing. In yet another strategy, the antisense nucleic acid is incorporated in a ribozyme capable of specifically cleaving the target mRNA.

Example 19

10 Preparation and Use of Antisense Oligonucleotides

The antisense nucleic acid molecules to be used in gene therapy is either DNA or RNA sequences. They may comprise a sequence complementary to the sequence of the PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, or a nucleic acid encoding a PG1 protein responsible for a detectable phenoytpe. The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex having sufficient stability to inhibit the expression of the PG1 mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green et al., Ann. Rev. Biochem. 55:569-597 (1986) and Izant and Weintraub, Cell 36:1007-1015 (1984).

In some strategies, antisense molecules are obtained by reversing the orientation of the PG1 coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules is transcribed using in vitro transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript. Another approach involves transcription of PG1 antisense nucleic acids in vivo by operably linking DNA containing the antisense sequence to a promoter in an expression vector.

Alternatively, oligonucleotides which are complementary to the strand of the PG1 gene normally transcribed in the cell is synthesized in vitro. Thus, the antisense PG1 nucleic acids are complementary to the PG1 mRNA and are capable of hybridizing to the mRNA to create a duplex. In some embodiments, the PG1 antisense sequences may contain modified sugar phosphate backbones to increase stability and make them less sensitive to RNase activity. Examples of modifications suitable for use in antisense strategies are described by Rossi et al., Pharmacol. Ther. 50(2):245-254, (1991).

Various types of antisense oligonucleotides complementary to the sequence of the PG1 genomic DNA of SEQ ID NO: 179, the PG1 cDNA of SEQ ID NO: 3, or a nucleic acid encoding a PG1 protein responsible for a detectable phenoytpe is used. In one preferred embodiment, stable and semi-stable antisense oligonucleotides as described in International Application No. PCT WO94/23026, are used to inhibit the expression of the PG1 gene. In these molecules, the 3• end or both the 3• and 5• ends are engaged in intramolecular hydrogen bonding between complementary base pairs. These molecules are

35

better able to withstand exonuclease attacks and exhibit increased stability compared to conventional antisense oligonucleotides.

In another preferred embodiment, the antisense oligodeoxynucleotides described in International Application No. WO 95/04141, are used to inhibit expression of the PG1 gene.

In yet another preferred embodiment, the covalently cross-linked antisense oligonucleotides described in International Application No. WO 96/31523, are used to inhibit expression of the PG1 gene. These double- or single-stranded oligonucleotides comprise one or more, respectively, inter- or intra-oligonucleotide covalent cross-linkages, wherein the linkage consists of an amide bond between a primary amine group of one strand and a carboxyl group of the other strand or of the same strand, respectively, the primary amine group being directly substituted in the 2' position of the strand nucleotide monosaccharide ring, and the carboxyl group being carried by an aliphatic spacer group substituted on a nucleotide or nucleotide analog of the other strand or the same strand, respectively.

The antisense oligodeoxynucleotides and oligonucleotides disclosed in International Application No. WO 92/18522, may also be used to inhibit the expression of the PG1 gene. These molecules are stable to degradation and contain at least one transcription control recognition sequence which binds to control proteins and are effective as decoys therefor. These molecules may contain "hairpin" structures, "dumbbell" structures, "modified dumbbell" structures, "cross-linked" decoy structures and "loop" structures.

In another preferred embodiment, the cyclic double-stranded oligonucleotides described in 20 European Patent Application No. 0 572 287 A2, are used to inhibit the expression of the PG1 gene. These ligated oligonucleotide "dumbbells" contain the binding site for a transcription factor which binds to the PG1 promoter and inhibits expression of the gene under control of the transcription factor by sequestering the factor.

Use of the closed antisense oligonucleotides disclosed in International Application No. WO 92/19732, is also contemplated. Because these molecules have no free ends, they are more resistant to degradation by exonucleases than are conventional oligonucleotides. These oligonucleotides is multifunctional, interacting with several regions which are not adjacent to the target mRNA.

The appropriate level of antisense nucleic acids required to inhibit PG1 gene expression is determined using in vitro expression analysis. The antisense molecule is introduced into the cells by diffusion, injection, infection or transfection using procedures known in the art. For example, the antisense nucleic acids can be introduced into the body as a bare or naked oligonucleotide, oligonucleotide encapsulated in lipid, oligonucleotide sequence encapsidated by viral protein, or as an oligonucleotide operably linked to a promoter contained in an expression vector. The expression vector is any of a variety of expression vectors known in the art, including retroviral or viral vectors, vectors capable of extrachromosomal replication, or integrating vectors. The vectors is DNA or RNA.

PCT/IB98/02133 WO 99/32644 101

The PG1 antisense molecules are introduced onto cell samples at a number of different concentrations preferably between 1x10⁻¹⁰M to 1x10⁻⁴M. Once the minimum concentration that can adequately control gene expression is identified, the optimized dose is translated into a dosage suitable for use in vivo. For example, an inhibiting concentration in culture of 1x10⁻⁷ translates into a dose of approximately 0.6 mg/kg bodyweight. Levels of oligonucleotide approaching 100 mg/kg bodyweight or higher is possible after testing the toxicity of the oligonucleotide in laboratory animals. It is additionally contemplated that cells from the vertebrate are removed, treated with the antisense oligonucleotide, and reintroduced into the vertebrate.

It is further contemplated that the PG1 antisense oligonucleotide sequence is incorporated into a ribozyme sequence to enable the antisense to specifically bind and cleave its target mRNA. For technical applications of ribozyme and antisense oligonucleotides see Rossi et al., supra.

In a preferred application of this invention, antibody-mediated tests such as RIAs and ELISA, functional assays, or radiolabeling are used to determine the effectiveness of antisense inhibition on PG1 expression.

The PG1 cDNA, the PG1 genomic DNA, and the PG1 alleles of the present invention may also be used in gene therapy approaches based on intracellular triple helix formation. Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity as it is associated with a particular gene. The PG1 cDNA, PG1 genomic DNA, or PG1 allele of the present invention or, more preferably, a portion of those sequences, can be used to inhibit gene expression in individuals suffering from prostate cancer or another detectable phenotype or individuals at risk for developing prostate cancer or another detectable phenotype at a later date as a result of their PG1 genotype. Similarly, a portion of the PG1 cDNA, the PG1 genomic DNA, or the PG1 alleles can be used to study the effect of inhibiting PG1 transcription within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies, such as 25 those described in Example 20, below. However, homopyrimidine sequences can also inhibit gene oligonucleotides bind to the major homopyrimidine expression. Such homopurine:homopyrimidine sequences. Thus, both types of sequences from the PG1 cDNA, the PG1 genomic DNA, and the PG1 alleles are contemplated within the scope of this invention.

Example 20

The sequences of the PG1 cDNA, the PG1 genomic DNA, and the PG1 alleles are scanned to 30 identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting PG1 expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting PG1 expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which express the PG1 gene. The oligonucleotides is prepared on an oligonucleotide synthesizer or they is purchased

commercially from a company specializing in custom oligonucleotide synthesis, such as GENSET, Paris, France.

The oligonucleotides is introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

Treated cells are monitored for altered cell function or reduced PG1 expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the PG1 gene in cells which have been treated with the oligonucleotide.

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced in vivo using the techniques described above and in Example 19 at a dosage calculated based on the in vitro results, as described in Example 19.

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin et al. (Science 245:967-971 (1989).

Alternatively, the PG1 cDNA, the PG1 genomic DNA, and the PG1 alleles of the present invention is used in gene therapy approaches in which expression of the PG1 protein is beneficial, as described in Example 21 below.

20 <u>Example 21</u>

The PG1 cDNA, the PG1 genomic DNA, and the PG1 alleles of the present invention may also be used to express the PG1 protein or a portion thereof in a host organism to produce a beneficial effect. In such procedures, the PG1 protein is transiently expressed in the host organism or stably expressed in the host organism. The expressed PG1 protein is used to treat conditions resulting from a lack of PG1 expression or conditions in which augmentation of existing levels of PG1 expression is beneficial.

A nucleic acid encoding the PG1 proteins of SEQ ID NO: 4, SEQ ID NO:5, or a PG1 allele is introduced into the host organism. The nucleic acid is introduced into the host organism using a variety of techniques known to those of skill in the art. For example, the nucleic acid is injected into the host organism as naked DNA such that the encoded PG1 protein is expressed in the host organism, thereby producing a beneficial effect.

Alternatively, the nucleic acid encoding the PG1 proteins of SEQ ID NO: 4, SEQ ID NO: 5, or a PG1 allele is cloned into an expression vector downstream of a promoter which is active in the host organism. The expression vector is any of the expression vectors designed for use in gene therapy, including viral or retroviral vectors.

The expression vector is directly introduced into the host organism such that the PG1 protein is expressed in the host organism to produce a beneficial effect. In another approach, the expression vector

WO 99/32644 PCT/IB98/02133

is introduced into cells in vitro. Cells containing the expression vector are thereafter selected and introduced into the host organism, where they express the PG1 protein to produce a beneficial effect.

IX. ISOLATION OF PG1 cDNA FROM NONHUMAN MAMMALS

The present invention encompasses mammalian PG1 sequences including genomic and cDNA sequences, as well as polypeptide sequences. The present invention also encompasses the use of PG1 genomic and cDNA sequences of the invention, including SEQ ID NOs: 179, 3, 182, and 183, in methods of isolating and characterizing PG1 nucleotide sequences derived from nonhuman mammals, in addition to sequences derived from human sequences. The human and mouse PG1 nucleic acid sequences of the invention can be used to construct primers and probes for amplifying and identifying PG1 genes in other nonhuman animals particularly mammals. The primers and probes used to identify nonhuman PG1 sequences is selected and used for the isolation of nonhuman PG1 utilizing the same techniques described above in Examples 4, 5, 6, 12 and 13.

In addition, sequence analysis of other homologous proteins is used to optimize the sequences of these primers and probes. As described above in the Analysis of the PG1 Protein Sequence, three boxes of homology were identified in the structure of the PG1 protein product when compared to proteins from a diverse range of organisms. See Figure 9. Using the assumption that the nucleotide sequences for these homologous proteins also show a high degree of homology, it is possible to construct primers that are specific for the PCR amplification of PG1 cDNA in nonhuman mammals.

Example 22

The primers BOXIed: AATCATCAAAGCACAGTTGACTGGAT (SEQ ID NO: 77) and BOXIIIer: ATAAACCACCGTAACATCATAAATTGCATCTAA (SEQ ID NO: 78) were designed as PCR primers from the human PG1 sequences after comparison with the sequence homologies of Figure 9. The BOXIed (SEQ ID NO: 77) and BOXIIIer (SEQ ID NO: 78) primers were used to amplify a mouse PG1 cDNA sequence from mouse liver marathon-ready cDNA (Clontech) under the conditions described above in Example 4. This PCR reaction yielded a product of approximately 400 base pairs, the boxI-boxIII fragment, which was subjected to automated dideoxy terminator sequencing and electrophoresed on ABI 377 sequencers as described above. Sequence analysis confirmed very high homology to human PG1 both at the nucleic acid and protein levels.

Primers were designed for RACE analysis using the 400 base pair boxI-boxIII fragment. Further sequence information was obtained using 5' and 3' RACE reactions on mouse liver marathon cDNA using two sets of these nested PCR primers: moPG1RACE5.350: AATCAAAAGCAACGTGAGTGGC (SEQ ID NO: 94) and moPG1RACE5.276: GCAAATGCCTGACTGGCTGA (SEQ ID NO: 93) for the 5' RACE reaction and moPG1RACE3.18: CTGCCAGACAGGATGCCCTA (SEQ ID NO: 90) and moPG1RACE3.63: ACAAGTTAAAATGGCTTCCGCTG (SEQ ID NO: 91) for the 3' RACE reaction.

The PCR products of the RACE reactions were sequenced by primer walking using the following primers:

	104	
moPGrace3S473:	GAGATAAAAG ATAGGTTGCT CA	(SEQ ID NO: 79);
moPGrace3S526:	AAGAAACAAA TTTCCTGGG	(SEQ ID NO: 80);
moPGrace3S597:	TCTTGGGGAG TTTGACTG	(SEQ ID NO: 81);
moPGrace5R323:	GACCCCGGTG TAGTTCTC	(SEQ ID NO: 82);
moPGrace5R372:	CAGTAAAGCC GGTCGTC	(SEQ ID NO: 83);
moPGrace5R444:	CAGGCCAGCA GGTAGGT	(SEQ ID NO: 84);
moPGrace5R492:	AGCAGGTAGC GCATAGAGT	(SEQ ID NO: 85).

WO 99/32644

5

PCT/IB98/02133

Again a high degree of homology between the mouse sequence obtained from the primer walking and the human PG1 sequence was observed. An additional pair of nested primers were designed and utilized to further extend the 3' mouse PG1 sequence in yet another RACE reaction, moPG3RACE2: TGGGCACCTG GTTGTATGGA (SEQ ID NO: 95) and moPG3RACE2n: TCCTTGGCTG CCTGTGGTTT (SEQ ID NO:96). The PCR product of this final RACE reaction was also sequenced by primer walking using the following primers:

	moPG1RACE3R94:	CAAATGCATG TTGGCTGT	(SEQ ID NO: 92);
	moPG3RACES20:	GATGGCTACA CATTGTATCA C	(SEQ ID NO: 97);
	moPG3RACES5:	TCCTGAATTA AATAAGGAGT TTTC	(SEQ ID NO: 98);
	moPG3RACES90:	GTTTGTTATT AAAGCATAAG CAAG	(SEQ ID NO: 99).

The overlap in the 5' RACE, boxI-boxIII, and 3' RACE fragments allowed a single contiguous coding sequence for the mouse PG1 ortholog to be generated alignment of the three fragments. Primers 20 were chosen from near the 5' and 3' ends of this predicted contiguous sequence (contig) in order to confirm the existence of such a transcript. PCR amplification was performed again on mouse liver marathon-ready cDNA (Clontech) with the chosen primers, moPG15: TGGCGAGCCGAGAGGATG (SEQ ID NO: 87) and moPG13LR2: GGAAACAATGTGATACAATGTGTAGCC (SEQ ID NO: 86) under the PCR conditions described above in Example 4. The resulting PCR product was a roughly 1.2 25 kb DNA molecule and was shown to have an identical sequence to that of the deduced contig. Finally modified versions of the moPG15 and moPG13LR2 primers with the addition of EcoRI and BamHI sites, moPG15EcoRI: CGTGAATTCTGGCGAGCCGAGAGGATG (SEQ ID NO: 89) moPG15Bam1: CGTGGATCCGGAAACAATGTGATACAATGTGTAGCC (SEQ ID NO: 88) were used to obtain a PCR product that could be cloned into a pSKBluescript plasmid (Stratagene) cleaved with EcoRI and BamHI restriction enzymes. The mouse PG1 cDNA in the resulting construct was subjected to automated dideoxy terminator sequencing and electrophoresed on ABI 377 sequencers as described above. The sequence for mouse PG1 cDNA is reported in SEQ ID NO: 72, and the deduced amino acid sequence corresponding to the cDNA is reported in SEQ ID NO: 74.

Example 23

A mouse BAC library was constructed by the cloning of BamHI partially digested DNA of pluripotent embryonic stem cells, cell line ES-E14TG2a (ATCC CRL-1821) into pBeloBACII vector

plasmid. Approximately fifty-six thousand clones with an average inset size of 120 kb were picked individually and pooled for PCR screening as described above for human BAC library screening. These pools were screened with STS g34292 derived from the region of the mouse PG1 transcript corresponding to exon6 of the human gene. The upstream and downstream primers defining this STS are: upstream amplification primer for g34292: ATTAAAACAC GTACTGACAC CA (SEQ ID NO: 75), and downstream amplification primer for g34292: AGTCATGGAT GGTGGATTT (SEQ ID NO: 76). BAC C0281H06 tested positive for hybridizing to g34292. This BAC was isolated and sequenced by sub-cloning into pGenDel sequencing vector. The resulting partial genomic sequence for mouse PG1 is reported in SEQ ID NO: 73. This process was repeated and the resulting partial genomic sequences for mouse PG1 is reported in SEQ ID NO: 182 and 183.

Other mammalian PG1 cDNA and genomic sequences can be isolated by the methods of the present invention. PG1 genes in mammalian species have a region of at least 100, preferably 200, more preferably 500 nucleotides in each mammal's most abundant transcription species which has at least 75%, preferably 85%, more preferably 95% sequence homology to the most abundant human or mouse cDNA species (SEQ ID NO: 3). PG1 proteins in mammalian species have a region of at least 40, preferably 90, more preferably 160 amino acids in the deduced amino acid sequence of the most abundant PG1 transcirption species which has at least 75%, preferably 85%, more preferably 95% sequence homology to the deduced amino acid sequence of the most abundant human or mouse translations species (SEQ ID NO: 4 or 74).

20 X. METHODS FOR GENOTYPING AN INDIVIDUAL FOR BIALLELIC MARKERS

Methods are provided to genotype a biological sample for one or more biallelic markers of the present invention, all of which is performed in vitro. Such methods of genotyping comprise determining the identity of a nucleotide at an PG1-related biallelic marker by any method known in the art. These methods find use in genotyping case-control populations in association studies as well as individuals in the context of detection of alleles of biallelic markers which, are known to be associated with a given trait, in which case both copies of the biallelic marker present in individual's genome are determined so that an individual is classified as homozygous or heterozygous for a particular allele.

These genotyping methods can be performed nucleic acid samples derived from a single individual or pooled DNA samples.

Genotyping can be performed using similar methods as those described above for the identification of the biallelic markers, or using other genotyping methods such as those further described below. In preferred embodiments, the comparison of sequences of amplified genomic fragments from different individuals is used to identify new biallelic markers whereas microsequencing is used for genotyping known biallelic markers in diagnostic and association study applications.

X.A. Source of DNA for genotyping

Any source of nucleic acids, in purified or non-purified form, can be utilized as the starting nucleic acid, provided it contains or is suspected of containing the specific nucleic acid sequence desired. DNA or RNA is extracted from cells, tissues, body fluids. As for the source of genomic 5 DNA to be subjected to analysis, any test sample can be foreseen without any particular limitation. These test samples include biological samples, which can be tested by the methods of the present invention described herein, and include human and animal body fluids such as whole blood, serum, plasma, cerebrospinal fluid, urine, lymph fluids, and various external secretions of the respiratory, intestinal and genitourinary tracts, tears, saliva, milk, white blood cells, myelomas and the like; 10 biological fluids such as cell culture supernatants; fixed tissue specimens including tumor and nontumor tissue and lymph node tissues; bone marrow aspirates and fixed cell specimens. The preferred source of genomic DNA used in the present invention is from peripheral venous blood of each donor. Techniques to prepare genomic DNA from biological samples are well known to the skilled technician. While nucleic acids for use in the genotyping methods of the invention can be derived from any mammalian source, the test subjects and individuals from which nucleic acid samples are 15 taken are generally understood to be human.

X.B. Amplification Of DNA Fragments Comprising Biallelic Markers

Methods and polynucleotides are provided to amplify a segment of nucleotides comprising one or more biallelic marker of the present invention. It will be appreciated that amplification of DNA fragments comprising biallelic markers is used in various methods and for various purposes and is not restricted to genotyping. Nevertheless, many genotyping methods, although not all, require the previous amplification of the DNA region carrying the biallelic marker of interest. Such methods specifically increase the concentration or total number of sequences that span the biallelic marker or include that site and sequences located either distal or proximal to it. Diagnostic assays may also rely on amplification of DNA segments carrying a biallelic marker of the present invention.

Amplification of DNA is achieved by any method known in the art. The established PCR (polymerase chain reaction) method or by developments thereof or alternatives. Amplification methods which can be utilized herein include but are not limited to Ligase Chain Reaction (LCR) as described in EP A 320 308 and EP A 439 182, Gap LCR (Wolcott, M.J., Clin. Mcrobiol. Rev. 5:370-386), the so-called "NASBA" or "3SR" technique described in Guatelli J.C. et al. (*Proc. Natl. Acad. Sci. USA* 87:1874-1878, 1990) and in Compton J. (*Nature* 350:91-92, 1991), Q-beta amplification as described in European Patent Application no 4544610, strand displacement amplification as described in Walker et al. (*Clin. Chem.* 42:9-13, 1996) and EP A 684 315 and, target mediated amplification as described in PCT Publication WO 9322461.

LCR and Gap LCR are exponential amplification techniques, both depend on DNA ligase to join adjacent primers annealed to a DNA molecule. In Ligase Chain Reaction (LCR), probe pairs are

35

used which include two primary (first and second) and two secondary (third and fourth) probes, all of which are employed in molar excess to target. The first probe hybridizes to a first segment of the target strand and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate-3'hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused product. In addition, a third (secondary) probe can hybridize to a portion of the first probe and a fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Of course, if the target is initially double stranded, the secondary probes also will hybridize to the target complement in the first instance. Once the ligated strand of primary probes is separated from the target strand, it will hybridize with the third and fourth probes which can be ligated to form a complementary, secondary ligated product. It is important to realize that the ligated products are functionally equivalent to either the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. A method for multiplex LCR has also been described (WO 9320227). Gap LCR (GLCR) is a version of LCR where the probes are not adjacent but are separated by 2 to 3 bases.

For amplification of mRNAs, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Patent No. 5,322,770 or, to use Asymmetric Gap LCR (RT-AGLCR) as described by Marshall R.L. et al. (PCR Methods and Applications 4:80-84, 1994).

20 AGLCR is a modification of GLCR that allows the amplification of RNA.

Some of these amplification methods are particularly suited for the detection of single nucleotide polymorphisms and allow the simultaneous amplification of a target sequence and the identification of the polymorphic nucleotide as it is further described in X.C.

The PCR technology is the preferred amplification technique used in the present invention. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see Molecular Cloning to Genetic Engineering White, B.A. Ed. in *Methods in Molecular Biology* 67: Humana Press, Totowa (1997) and the publication entitled "PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press). In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites. PCR has further been described in several patents including US Patents 4,683,195, 4,683,202 and 4,965,188.

The identification of biallelic markers as described above allows the design of appropriate oligonucleotides, which can be used as primers to amplify DNA fragments comprising the biallelic markers of the present invention. Amplification can be performed using the primers initially used to discover new biallelic markers which are described herein or any set of primers allowing the amplification of a DNA fragment comprising a biallelic marker of the present invention. Primers can be prepared by any suitable method. As for example, direct chemical synthesis by a method such as the phosphodiester method of Narang S.A. et al. (Methods Enzymol. 68:90-98, 1979), the phosphodiester method of Brown E.L. et al. (Methods Enzymol. 68:109-151, 1979), the diethylphosphoramidite method of Beaucage et al. (Tetrahedron Lett. 22:1859-1862, 1981) and the solid support method described in EP 0 707 592.

In some embodiments the present invention provides primers for amplifying a DNA fragment containing one or more biallelic markers of the present invention. It will be appreciated that the amplification primers listed in the present specification are merely exemplary and that any other set of primers which produce amplification products containing one or more biallelic markers of the present invention.

The primers are selected to be substantially complementary to the different strands of each specific sequence to be amplified. The length of the primers of the present invention can range from 8 to 100 nucleotides, preferably from 8 to 50, 8 to 30 or more preferably 8 to 25 nucleotides. Shorter primers tend to lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer primers are expensive to produce and can sometimes self-hybridize to form hairpin structures. The formation of stable hybrids depends on the melting temperature TM of the DNA. The Tm depends on the length of the primer, the ionic strength of the solution and the G+C content. The higher the G+C content of the primer, the higher is the melting temperature because G:C pairs are held by three H bonds whereas A:T pairs have only two. The G+C content of the amplification primers of the present invention preferably ranges between 10 and 75 %, more preferably between 35 and 60 %, and most preferably between 40 and 55 %. The appropriate length for primers under a particular set of assay conditions is empirically determined by one of skill in the art.

The spacing of the primers determines the length of the segment to be amplified. In the context of the present invention amplified segments carrying biallelic markers can range in size from at least about 25 bp to 35 kbp. Amplification fragments from 25-3000 bp are typical, fragments from 50-1000 bp are preferred and fragments from 100-600 bp are highly preferred. It will be appreciated that amplification primers for the biallelic markers is any sequence which allow the specific amplification of any DNA fragment carrying the markers. Amplification primers is labeled or immobilized on a solid support as described in Section II.

X.C. Methods of Genotyping DNA samples for Biallelic Markers

Any method known in the art can be used to identify the nucleotide present at a biallelic marker site. Since the biallelic marker allele to be detected has been identified and specified in the present invention, detection will prove routine for one of ordinary skill in the art by employing any of a number of techniques. Many genotyping methods require the previous amplification of the DNA 5 region carrying the biallelic marker of interest. While the amplification of target or signal is often preferred at present, ultrasensitive detection methods which do not require amplification are also encompassed by the present genotyping methods. Methods well-known to those skilled in the art that can be used to detect biallelic polymorphisms include methods such as, conventional dot blot analyzes, single strand conformational polymorphism analysis (SSCP) described by Orita et al. (Proc. 10 Natl. Acad. Sci. U.S.A 86:27776-2770, 1989), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis, mismatch cleavage detection, and other conventional techniques as described in Sheffield, V.C. et al. (Proc. Natl. Acad. Sci. USA 49:699-706, 1991), White et al. (Genomics 12:301-306, 1992), Grompe, M. et al. (Proc. Natl. Acad. Sci. USA 86:5855-5892, 1989) and Grompe, M. (Nature Genetics 5:111-117, 1993). Another method for determining the identity of the nucleotide present at a particular polymorphic site employs a specialized exonuclease-resistant nucleotide derivative as described in US patent 4,656,127.

Preferred methods involve directly determining the identity of the nucleotide present at a biallelic marker site by sequencing assay, allele-specific amplification assay, or hybridization assay. The following is a description of some preferred methods. A highly preferred method is the microsequencing technique. The term "sequencing assay" is used herein to refer to polymerase extension of duplex primer/template complexes and includes both traditional sequencing and microsequencing.

1) Sequencing assays

The nucleotide present at a polymorphic site can be determined by sequencing methods. In a preferred embodiment, DNA samples are subjected to PCR amplification before sequencing as described above. Methods for sequencing DNA using either the dideoxy-mediated method (Sanger method) or the Maxam-Gilbert method are widely known to those of ordinary skill in the art. Such methods are for example disclosed in Maniatis et al. (Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Second Edition, 1989). Alternative approaches include hybridization to high-density DNA probe arrays as described in Chee et al. (Science 274, 610, 1996).

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. The products of the sequencing reactions are run on sequencing gels and the sequences are determined using gel image analysis.

The polymorphism detection in a pooled sample is based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position.

Because each dideoxy terminator is labeled with a different fluorescent molecule, the two peaks

corresponding to a biallelic site present distinct colors corresponding to two different nucleotides at the same position on the sequence. However, the presence of two peaks can be an artifact due to background noise. To exclude such an artifact, the two DNA strands are sequenced and a comparison between the peaks is carried out. In order to be registered as a polymorphic sequence, the polymorphism has to be detected on both strands.

The above procedure permits those amplification products, which contain biallelic markers to be identified. The detection limit for the frequency of biallelic polymorphisms detected by sequencing pools of 100 individuals is approximately 0.1 for the minor allele, as verified by sequencing pools of known allelic frequencies.

Microsequencing assays

10

In microsequencing methods, the nucleotide at a polymorphic site in a target DNA is detected by a single nucleotide primer extension reaction. This method involves appropriate microsequencing primers which, hybridize just upstream of the polymorphic base of interest in the target nucleic acid. A polymerase is used to specifically extend the 3' end of the primer with one single ddNTP (chain terminator) complementary to the nucleotide at the polymorphic site. Next the identity of the incorporated nucleotide is determined in any suitable way.

Typically, microsequencing reactions are carried out using fluorescent ddNTPs and the extended microsequencing primers are analyzed by electrophoresis on ABI 377 sequencing machines to determine the identity of the incorporated nucleotide as described in EP 412 883. Alternatively capillary electrophoresis can be used in order to process a higher number of assays simultaneously.

Different approaches can be used to detect the nucleotide added to the microsequencing primer. A homogeneous phase detection method based on fluorescence resonance energy transfer has been described by Chen and Kwok (Nucleic Acids Research 25:347-353 1997) and Chen et al. (Proc. Natl. Acad. Sci. USA 94/20 10756-10761,1997). In this method amplified genomic DNA fragments containing polymorphic sites are incubated with a 5'-fluorescein-labeled primer in the presence of allelic dye-labeled dideoxyribonucleoside triphosphates and a modified Taq polymerase. The dye-labeled primer is extended one base by the dye-terminator specific for the allele present on the template. At the end of the genotyping reaction, the fluorescence intensities of the two dyes in the reaction mixture are analyzed directly without separation or purification. All these steps can be performed in the same tube and the fluorescence changes can be monitored in real time. Alternatively, the extended primer is analyzed by MALDI-TOF Mass Spectrometry. The base at the polymorphic site is identified by the mass added onto the microsequencing primer (see Haff L.A. and Smirnov I.P., Genome Research, 7:378-388, 1997).

Microsequencing is achieved by the established microsequencing method or by developments or derivatives thereof. Alternative methods include several solid-phase microsequencing techniques. The basic microsequencing protocol is the same as described previously, except that the method is

conducted as a heterogenous phase assay, in which the primer or the target molecule is immobilized or captured onto a solid support. To simplify the primer separation and the terminal nucleotide addition analysis, oligonucleotides are attached to solid supports or are modified in such ways that permit affinity separation as well as polymerase extension. The 5' ends and internal nucleotides of synthetic oligonucleotides can be modified in a number of different ways to permit different affinity separation approaches, e.g., biotinylation. If a single affinity group is used on the oligonucleotides, the oligonucleotides can be separated from the incorporated terminator regent. This eliminates the need of physical or size separation. More than one oligonucleotide can be separated from the terminator reagent and analyzed simultaneously if more than one affinity group is used. This permits the analysis of several nucleic acid species or more nucleic acid sequence information per extension reaction. The affinity group need not be on the priming oligonucleotide but could alternatively be present on the template. For example, immobilization can be carried out via an interaction between biotinylated DNA and streptavidin-coated microtitration wells or avidin-coated polystyrene particles. In the same manner oligonucleotides or templates is attached to a solid support in a high-density 15 format. In such solid phase microsequencing reactions, incorporated ddNTPs can be radiolabeled (Syvänen, Clinica Chimica Acta 226:225-236, 1994) or linked to fluorescein (Livak and Hainer, Human Mutation 3:379-385,1994). The detection of radiolabeled ddNTPs can be achieved through scintillation-based techniques. The detection of fluorescein-linked ddNTPs can be based on the binding of antifluorescein antibody conjugated with alkaline phosphatase, followed by incubation 20 with a chromogenic substrate (such as p-nitrophenyl phosphate). Other possible reporter-detection pairs include: ddNTP linked to dinitrophenyl (DNP) and anti-DNP alkaline phosphatase conjugate (Harju et al., Clin. Chem. 39/11 2282-2287, 1993) or biotinylated ddNTP and horseradish peroxidaseconjugated streptavidin with o-phenylenediamine as a substrate (WO 92/15712). As yet another alternative solid-phase microsequencing procedure, Nyren et al. (Analytical Biochemistry 208:171-25 175, 1993) described a method relying on the detection of DNA polymerase activity by an enzymatic luminometric inorganic pyrophosphate detection assay (ELIDA).

Pastinen et al. (Genome research 7:606-614, 1997) describe a method for multiplex detection of single nucleotide polymorphism in which the solid phase minisequencing principle is applied to an oligonucleotide array format. High-density arrays of DNA probes attached to a solid support (DNA chips) are further described in X.C.5.

In one aspect the present invention provides polynucleotides and methods to genotype one or more biallelic markers of the present invention by performing a microsequencing assay. It will be appreciated that any primer having a 3' end immediately adjacent to the polymorphic nucleotide is used. However, polynucleotides comprising at least 8, 12, 15, 20, 25, or 30 consecutive nucleotides of the sequence immediately adjacent to the biallelic marker and having a 3' terminus immediately

upstream of the corresponding biallelic marker are well suited for determining the identity of a nucleotide at biallelic marker site.

Similarly, it will be appreciated that microsequencing analysis is performed for any biallelic marker or any combination of biallelic markers of the present invention.

Mismatch detection assays based on polymerases and ligases

In one aspect the present invention provides polynucleotides and methods to determine the allele of one or more biallelic markers of the present invention in a biological sample, by mismatch detection assays based on polymerases and/or ligases. These assays are based on the specificity of polymerases and ligases. Polymerization reactions places particularly stringent requirements on correct base pairing of the 3' end of the amplification primer and the joining of two oligonucleotides hybridized to a target DNA sequence is quite sensitive to mismatches close to the ligation site, especially at the 3' end. Methods, primers and various parameters to amplify DNA fragments comprising biallelic markers of the present invention are further described above in X.B.

Allele specific amplification

10

15

Discrimination between the two alleles of a biallelic marker can also be achieved by allele specific amplification, a selective strategy, whereby one of the alleles is amplified without amplification of the other allele. This is accomplished by placing the polymorphic base at the 3' end of one of the amplification primers. Because the extension forms from the 3'end of the primer, a mismatch at or near this position has an inhibitory effect on amplification. Therefore, under 20 appropriate amplification conditions, these primers only direct amplification on their complementary allele. Designing the appropriate allele-specific primer and the corresponding assay conditions are well with the ordinary skill in the art.

Ligation/amplification based methods

The "Oligonucleotide Ligation Assay" (OLA) uses two oligonucleotides which are designed 25 to be capable of hybridizing to abutting sequences of a single strand of a target molecules. One of the oligonucleotides is biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut, and create a ligation substrate that can be captured and detected. OLA is capable of detecting single nucleotide polymorphisms and is advantageously combined with PCR as described by 30 Nickerson D.A. et al. (Proc. Natl. Acad. Sci. U.S.A. 87:8923-8927, 1990). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

Other methods which are particularly suited for the detection of single nucleotide polymorphism include LCR (ligase chain reaction), Gap LCR (GLCR) which are described above in X.B. As mentioned above LCR uses two pairs of probes to exponentially amplify a specific target. 35 The sequences of each pair of oligonucleotides, is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such hybridization forms a substrate for a template-

dependant ligase. In accordance with the present invention, LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a biallelic marker site. In one embodiment, either oligonucleotide will be designed to include the biallelic marker site. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide that is complementary to the biallelic marker on the oligonucleotide. In an alternative embodiment, the oligonucleotides will not include the biallelic marker, such that when they hybridize to the target molecule, a "gap" is created as described in WO 90/01069. This gap is then "filled" with complementary dNTPs (as mediated by DNA polymerase), or by an additional pair of oligonucleotides. Thus at the end of each cycle, each single strand has a complement capable of serving as a target during the next cycle and exponential allele-specific amplification of the desired sequence is obtained.

Ligase/Polymerase-mediated Genetic Bit AnalysisTM is another method for determining the identity of a nucleotide at a preselected site in a nucleic acid molecule (WO 95/21271). This method 15 involves the incorporation of a nucleoside triphosphate that is complementary to the nucleotide present at the preselected site onto the terminus of a primer molecule, and their subsequent ligation to a second oligonucleotide. The reaction is monitored by detecting a specific label attached to the reaction's solid phase or by detection in solution.

2) Hybridization assay methods

A preferred method of determining the identity of the nucleotide present at a biallelic marker site involves nucleic acid hybridization. The hybridization probes, which can be conveniently used in such reactions, preferably include the probes defined herein. Any hybridization assay is used including Southern hybridization, Northern hybridization, dot blot hybridization and solid-phase hybridization (see Sambrook et al., Molecular Cloning - A Laboratory Manual, Second Edition, Cold 25 Spring Harbor Press, N.Y., 1989).

Hybridization refers to the formation of a duplex structure by two single stranded nucleic acids due to complementary base pairing. Hybridization can occur between exactly complementary nucleic acid strands or between nucleic acid strands that contain minor regions of mismatch. Specific probes can be designed that hybridize to one form of a biallelic marker and not to the other and therefore are able to discriminate between different allelic forms. Allele-specific probes are often used in pairs, one member of a pair showing perfect match to a target sequence containing the original allele and the other showing a perfect match to the target sequence containing the alternative allele. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Stringent, sequence specific hybridization conditions, under which a probe will hybridize only to the exactly complementary target sequence are well known

in the art (Sambrook et al., Molecular Cloning - A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., 1989). Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point TM for the specific sequence at a defined ionic strength and pH. By way of 5 example and not limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C, the preferred hybridization temperature, in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10⁶ cpm of ³²P-labeled probe. Alternatively, the hybridization step can be performed at 65°C in the presence of SSC buffer, 1 x SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1X SSC at 50°C for 45 min. Alternatively, filter washes can be performed in a solution containing 2 x SSC and 0.1% SDS, 15 or 0.5 x SSC and 0.1% SDS, or 0.1 x SSC and 0.1% SDS at 68°C for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. By way of example and not limitation, procedures using conditions of intermediate stringency are as follows: Filters containing DNA are prehybridized, and then hybridized at a temperature of 60°C in the presence of a 5 x SSC buffer and labeled probe. Subsequently, filters washes are performed in a solution containing 2x SSC 20 at 50°C and the hybridized probes are detectable by autoradiography. Other conditions of high and intermediate stringency which is used are well known in the art and as cited in Sambrook et al. (Molecular Cloning - A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., 1989) and Ausubel et al. (Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y., 1989).

Although such hybridizations can be performed in solution, it is preferred to employ a solid-phase hybridization assay. The target DNA comprising a biallelic marker of the present invention is amplified prior to the hybridization reaction. The presence of a specific allele in the sample is determined by detecting the presence or the absence of stable hybrid duplexes formed between the probe and the target DNA. The detection of hybrid duplexes can be carried out by a number of methods. Various detection assay formats are well known which utilize detectable labels bound to either the target or the probe to enable detection of the hybrid duplexes. Typically, hybridization duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Those skilled in the art will recognize that wash steps is employed to wash away excess target DNA or probe. Standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the primers and probes.

Two recently developed assays allow hybridization-based allele discrimination with no need for separations or washes (see Landegren U. et al., Genome Research, 8:769-776,1998). The TaqMan assay takes advantage of the 5' nuclease activity of Taq DNA polymerase to digest a DNA probe annealed specifically to the accumulating amplification product. TaqMan probes are labeled with a donor-acceptor dye pair that interacts via fluorescence energy transfer. Cleavage of the TaqMan probe by the advancing polymerase during amplification dissociates the donor dye from the quenching acceptor dye, greatly increasing the donor fluorescence. All reagents necessary to detect two allelic variants can be assembled at the beginning of the reaction and the results are monitored in real time (see Livak et al., Nature Genetics, 9:341-342, 1995). In an alternative homogeneous hybridization 10 based procedure, molecular beacons are used for allele discriminations. Molecular beacons are hairpin-shaped oligonucleotide probes that report the presence of specific nucleic acids in When they bind to their targets they undergo a conformational homogeneous solutions. reorganization that restores the fluorescence of an internally quenched fluorophore (Tyagi et al., Nature Biotechnology, 16:49-53, 1998).

The polynucleotides provided herein can be used in hybridization assays for the detection of biallelic marker alleles in biological samples. These probes are characterized in that they preferably comprise between 8 and 50 nucleotides, and in that they are sufficiently complementary to a sequence comprising a biallelic marker of the present invention to hybridize thereto and preferably sufficiently specific to be able to discriminate the targeted sequence for only one nucleotide variation. The GC 20 content in the probes of the invention usually ranges between 10 and 75 %, preferably between 35 and 60 %, and more preferably between 40 and 55 %. The length of these probes can range from 10, 15, 20, or 30 to at least 100 nucleotides, preferably from 10 to 50, more preferably from 18 to 35 nucleotides. A particularly preferred probe is 25 nucleotides in length. Preferably the biallelic marker is within 4 nucleotides of the center of the polynucleotide probe. In particularly preferred probes the biallelic marker is at the center of said polynucleotide. Shorter probes may lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer probes are expensive to produce and can sometimes selfhybridize to form hairpin structures. Methods for the synthesis of oligonucleotide probes have been described above and can be applied to the probes of the present invention.

Preferably the probes of the present invention are labeled or immobilized on a solid support. Labels and solid supports are further described in II. Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702, morpholino analogs which are described in U.S. Patents Numbered 5,185,444; 5,034,506 and 5,142,047. The probe may have to be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by

30

modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group. Alternatively, the 3' hydroxyl group simply can be cleaved, replaced or modified, U.S. Patent Application Serial No. 07/049,061 filed April 19, 1993 describes modifications, which can be used to render a probe non-extendable.

The probes of the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA or Northern hybridization to mRNA. The probes can also be used to detect PCR amplification products. By assaying the hybridization to an allele specific probe, one can detect the presence or absence of a biallelic marker allele in a given sample.

High-Throughput parallel hybridizations in array format are specifically encompassed within "hybridization assays" and are described below.

Hybridization to addressable arrays of oligonucleotides

Hybridization assays based on oligonucleotide arrays rely on the differences in hybridization stability of short oligonucleotides to perfectly matched and mismatched target sequence variants. Efficient access to polymorphism information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (the chip) at selected positions. Each DNA chip can contain thousands to millions of individual synthetic DNA probes arranged in a grid-like pattern and miniaturized to the size of a dime.

The chip technology has already been applied with success in numerous cases. For example, the screening of mutations has been undertaken in the BRCA1 gene, in *S. cerevisiae* mutant strains, and in the protease gene of HIV-1 virus (Hacia et al., *Nature Genetics*, 14(4):441-447, 1996; Shoemaker et al., *Nature Genetics*, 14(4):450-456, 1996; Kozal et al., *Nature Medicine*, 2:753-759, 1996). Chips of various formats for use in detecting biallelic polymorphisms can be produced on a customized basis by Affymetrix (GeneChipTM), Hyseq (HyChip and HyGnostics), and Protogene Laboratories.

In general, these methods employ arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual which, target sequences include a polymorphic marker. EP785280 describes a tiling strategy for the detection of single nucleotide polymorphisms. Briefly, arrays may generally be "tiled" for a large number of specific polymorphisms. By "tiling" is generally meant the synthesis of a defined set of oligonucleotide probes which is made up of a sequence complementary to the target sequence of interest, as well as preselected variations of that sequence, e.g., substitution of one or more given positions with one or more members of the basis set of monomers, i.e. nucleotides. Tiling strategies are further described in PCT application No. WO 95/11995. In a particular aspect, arrays are tiled for a number of specific, identified biallelic marker sequences. In particular the array is tiled to include a number of detection blocks, each detection block being specific for a specific biallelic marker or a set of biallelic markers.

25

For example, a detection block is tiled to include a number of probes, which span the sequence segment that includes a specific polymorphism. To ensure probes that are complementary to each allele, the probes are synthesized in pairs differing at the biallelic marker. In addition to the probes differing at the polymorphic base, monosubstituted probes are also generally tiled within the detection block. These monosubstituted probes have bases at and up to a certain number of bases in either direction from the polymorphism, substituted with the remaining nucleotides (selected from A, T, G, C and U). Typically the probes in a tiled detection block will include substitutions of the sequence positions up to and including those that are 5 bases away from the biallelic marker. The monosubstituted probes provide internal controls for the tiled array, to distinguish actual hybridization from artefactual cross-hybridization. Upon completion of hybridization with the target sequence and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data from the scanned array is then analyzed to identify which allele or alleles of the biallelic marker are present in the sample. Hybridization and scanning is carried out as described in PCT application No. WO 92/10092 and WO 95/11995 and US patent No.

5) Integrated Systems

Another technique, which is used to analyze polymorphisms, includes multicomponent integrated systems, which miniaturize and compartmentalize processes such as PCR and capillary electrophoresis reactions in a single functional device. An example of such technique is disclosed in US patent 5,589,136, which describes the integration of PCR amplification and capillary electrophoresis in chips.

Integrated systems can be envisaged mainly when microfluidic systems are used. These systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip to create functional microscopic valves and pumps with no moving parts. Varying the voltage controls the liquid flow at intersections between the micro-machined channels and changes the liquid flow rate for pumping across different sections of the microchip.

For genotyping biallelic markers, the microfluidic system may integrate nucleic acid amplification, microsequencing, capillary electrophoresis and a detection method such as laser-induced fluorescence detection.

XI. METHODS OF GENETIC ANALYSIS USING THE BIALLELIC MARKERS OF THE PRESENT INVENTION

The methods available for the genetic analysis of complex traits fall into different categories (see Lander and Schork, *Science*, 265, 2037-2048, 1994). In general, the biallelic markers of the present invention find use in any method known in the art to demonstrate a statistically significant

correlation between a genotype and a phenotype. The biallelic markers is used in linkage analysis and in allele-sharing methods. Preferably, the biallelic markers of the present invention are used to identify genes associated with detectable traits using association studies, an approach which does not require the use of affected families and which permits the identification of genes associated with complex and sporadic traits.

The genetic analysis using the biallelic markers of the present invention is conducted on any scale. The whole set of biallelic markers of the present invention or any subset of biallelic markers of the present invention is used. In some embodiments, any additional set of genetic markers including a biallelic marker of the present invention is used. As mentioned above, it should be noted that the biallelic markers of the present invention is included in any complete or partial genetic map of the human genome. These different uses are specifically contemplated in the present invention and claims.

XI.A. Linkage Analysis

Until recently, the identification of genes linked with detectable traits has mainly relied on a 15 statistical approach called linkage analysis. Linkage analysis involves proposing a model to explain the inheritance pattern of phenotypes and genotypes observed in a pedigree. Linkage analysis is based upon establishing a correlation between the transmission of genetic markers and that of a specific trait throughout generations within a family. In this approach, all members of a series of affected families are genotyped with a few hundred markers, typically microsatellite markers, which 20 are distributed at an average density of one every 10 Mb. By comparing genotypes in all family members, one can attribute sets of alleles to parental haploid genomes (haplotyping or phase determination). The origin of recombined fragments is then determined in the offspring of all families. Those that co-segregate with the trait are tracked. After pooling data from all families, statistical methods are used to determine the likelihood that the marker and the trait are segregating 25 independently in all families. As a result of the statistical analysis, one or several regions having a high probability of harboring a gene linked to the trait are selected as candidates for further analysis. The result of linkage analysis is considered as significant (i.e. there is a high probability that the region contains a gene involved in a detectable trait) when the chance of independent segregation of the marker and the trait is lower than 1 in 1000 (expressed as a LOD score > 3). Generally, the length of the candidate region identified as having a LOD score of greater than 3 using linkage analysis is between 2 and 20Mb. Once a candidate region is identified as described above, analysis of recombinant individuals using additional markers allows further delineation of the candidate region. Linkage analysis studies have generally relied on the use of a maximum of 5,000 microsatellite markers, thus limiting the maximum theoretical attainable resolution of linkage analysis to about 600

35

kb on average.

PCT/IB98/02133 WO 99/32644 119

Linkage analysis has been successfully applied to map simple genetic traits that show clear Mendelian inheritance patterns and which have a high penetrance (i.e., the ratio between the number of trait positive carriers of allele a and the total number of a carriers in the population). About 100 pathological trait-causing genes were discovered using linkage analysis over the last 10 years. In 5 most of these cases, the majority of affected individuals had affected relatives and the detectable trait was rare in the general population (frequencies less than 0.1%). In about 10 cases, such as Alzheimer's Disease, breast cancer, and Type II diabetes, the detectable trait was more common but the allele associated with the detectable trait was rare in the affected population. Thus, the alleles associated with these traits were not responsible for the trait in all sporadic cases.

Linkage analysis suffers from a variety of drawbacks. First, linkage analysis is limited by its reliance on the choice of a genetic model suitable for each studied trait. Furthermore, as already mentioned, the resolution attainable using linkage analysis is limited, and complementary studies are required to refine the analysis of the typical 2Mb to 20Mb regions initially identified through linkage analysis. In addition, linkage analysis approaches have proven difficult when applied to complex genetic traits, such as those due to the combined action of multiple genes and/or environmental factors. In such cases, too large an effort and cost are needed to recruit the adequate number of affected families required for applying linkage analysis to these situations, as recently discussed by Risch, N. and Merikangas, K. (Science, 273:1516-1517, 1996). Finally, linkage analysis cannot be applied to the study of traits for which no large informative families are available. Typically, this will 20 be the case in any attempt to identify trait-causing alleles involved in sporadic cases, such as alleles associated with positive or negative responses to drug treatment.

XI.B. Allele-Sharing methods

Whereas linkage analysis involves proposing a model to explain the inheritance pattern of phenotypes and genotypes in a pedigree, allele-sharing methods are not based on constructing a model, but rather on rejecting a model (see Lander and Schork, Science, 265, 2037-2048, 1994). More specifically, one tries to prove that the inheritance pattern of a chromosomal region is not consistent with random Mendelian segregation by showing that affected relatives inherit identical copies of the region more often than expected by chance. Because allele-sharing methods are nonparametric (that is, assume no model for the inheritance of the trait), they tend to be more useful 30 for the analysis of complex traits than linkage analysis. Affected relatives should show excess allele sharing even in the presence of incomplete penetrance and polygenic inheritance. Allele-Sharing methods involve studying affected relatives in a pedigree to determine how often a particular copy of a chromosomal region is shared identical-by-descent (IBD), that is, is inherited from a common ancestor within the pedigree. The frequency of IBD sharing at a locus can then be compared with random expectation. Affected sib pair analysis is a well-known special case and is the simplest form of this method.

However, as allele-sharing methods analyze affected relatives, they tend to be of limited value in the genetic analysis of drug responses or in the analysis of side effects to treatments. This type of analysis is impractical in such cases due to the lack of availability of familial cases. In fact, the likelihood of having more than one individual in a family being exposed to the same drug at the same time is very low.

XI.C. Association Studies

The present invention comprises methods for identifying one or several genes among a set of candidate genes that are associated with a detectable trait using the biallelic markers of the present invention. In one embodiment the present invention comprises methods to detect an association between a biallelic marker allele or a biallelic marker haplotype and a trait. Further, the invention comprises methods to identify a trait causing allele in linkage disequilibrium with any biallelic marker allele of the present invention.

As described above, alternative approaches can be employed to perform association studies: genome-wide association studies, candidate region association studies and candidate gene association studies. In a preferred embodiment, the biallelic markers of the present invention are used to perform candidate gene association studies. The candidate gene analysis clearly provides a short-cut approach to the identification of genes and gene polymorphisms related to a particular trait when some information concerning the biology of the trait is available. Further, the biallelic markers of the present invention is incorporated in any map of genetic markers of the human genome in order to perform genome-wide association studies. Methods to generate a high-density map of biallelic markers has been described in US Provisional Patent application serial number 60/082,614. The biallelic markers of the present invention may further be incorporated in any map of a specific candidate region of the genome (a specific chromosome or a specific chromosomal region for example).

As mentioned above, association studies is conducted within the general population and are not limited to studies performed on related individuals in affected families. Linkage disequilibrium and association studies are extremely valuable as they permit the analysis of sporadic or multifactor traits. Moreover, association studies represent a powerful method for fine-scale mapping enabling much finer mapping of trait causing alleles than linkage studies. Studies based on pedigrees often only narrow the location of the trait causing allele. Association studies and Linkage Disequilibrium mapping methods using the biallelic markers of the present invention can therefore be used to refine the location of a trait causing allele in a candidate region identified by Linkage Analysis or by Allele-Sharing methods. Moreover, once a chromosome segment of interest has been identified, the presence of a candidate gene such as a candidate gene of the present invention, in the region of interest can provide a shortcut to the identification of the trait causing allele. Biallelic markers of the

present invention can be used to demonstrate that a candidate gene is associated with a trait. Such uses are specifically contemplated in the present invention and claims.

1) Case-control populations (inclusion criteria)

Association studies do not concern familial inheritance and do not involve the analysis of large family pedigrees but compare the prevalence of a particular genetic marker, or a set of markers, in case-control populations. They are case-control studies based on comparison of unrelated case (affected or trait positive) individuals and unrelated control (random or unaffected or trait negative) individuals. The control group is composed of individuals chosen randomly or of unaffected (trait negative) individuals, preferably the control group is composed of unaffected or trait negative individuals. Further, the control group is preferably both ethnically- and age-matched to the case population. In the following "trait positive population", "case population" and "affected population" are used interchangeably.

An important step in the dissection of complex traits using association studies is the choice of case-control populations (see Lander and Schork, Science, 265, 2037-2048, 1994). Narrowing the 15 definition of the disease and restricting the patient population to extreme phenotypes allows one to work with a trait that is more nearly Mendelian in its inheritance pattern and more likely to be homogeneous (patients suffer from the disease for the same genetic reasons). Therefore, a major step in the choice of case-control populations is the clinical definition of a given trait or phenotype. Four criteria are often useful: clinical phenotype, age at onset, family history and severity. Preferably, in 20 order to perform efficient and significant association studies, such as those described herein, the trait under study should preferably follow a bimodal distribution in the population under study, presenting two clear non-overlapping phenotypes (trait positive and trait negative). Nevertheless, even in the absence of such bimodal distribution (as may in fact be the case for more complex genetic traits), any genetic trait may still be analyzed by the association method proposed here by carefully selecting the individuals to be included in the trait posițive and trait negative phenotypic groups. The selection procedure involves selecting individuals at opposite ends of the non-bimodal phenotype spectra of the trait under study, so as to include in these trait positive and trait negative populations individuals which clearly represent extreme, preferably non-overlapping phenotypes. This is particularly useful for continuous or quantitative traits (such as blood pressure for example). Selection of individuals at extreme ends of the trait distribution increases the ability to analyze these complex traits. The definition of the inclusion criteria for the case-control populations is an important aspect of association studies. The selection of those drastically different but relatively uniform phenotypes enables efficient comparisons in association studies and the possible detection of marked differences at the genetic level, provided that the sample sizes of the populations under study are significant enough.

Preferably, case-control populations to be included in association studies such as those proposed in the present invention consist of phenotypically homogeneous populations of individuals each representing 100% of the corresponding phenotype if the trait distribution is bimodal. If the trait distribution is non-bimodal, trait positive and trait negative populations consist of phenotypically uniform populations of individuals representing each between 1 and 98%, preferably between 1 and 80%, more preferably between 1 and 50%, and more preferably between 1 and 30%, most preferably between 1 and 20% of the total population under study, and selected among individuals exhibiting non-overlapping phenotypes. In some embodiments, the trait positive and trait negative groups consist of individuals exhibiting the extreme phenotypes within the studied population. The clearer the difference between the two trait phenotypes, the greater the probability of detecting an association with biallelic markers.

In preferred embodiments, a first group of between 50 and 300 trait positive individuals, preferably about 100 individuals, are recruited according to their phenotypes. A similar number of trait negative individuals are included in such studies.

In the present invention, typical examples of inclusion criteria include a diagnosis of cancer or prostate cancer or the evaluation of the response to anti-cancer or anti-prostate cancer agent or side effects to treatment with anti-cancer or anti-prostate cancer agents.

Suitable examples of association studies using biallelic markers including the biallelic markers of the present invention, are studies involving the following populations:

20 a case population suffering from a form of cancer and a healthy unaffected control population, or a case population suffering from a form of prostate cancer and a healthy unaffected control population, or

a case population treated with anticancer agents suffering from side-effects resulting from the treatment and a control population treated with the same agents showing no side-effects, or

a case population treated with anti-prostate cancer agents suffering from side-effects resulting from the treatment and a control population treated with the same agents showing no side-effects, or

a case population treated with anti-cancer agents showing a beneficial response and a control population treated with same agents showing no beneficial response, or

a case population treated with anti-prostate cancer agents showing a beneficial response and a control population treated with same agents showing no beneficial response.

2) Determining the frequency of an allele in case-control populations

Allelic frequencies of the biallelic markers in each of the populations can be determined using one of the methods described above under the in Section X. under the heading "Methods for genotyping an individual for biallelic markers", or any genotyping procedure suitable for this intended purpose. The frequency of a biallelic marker allele in a population can be determined by genotyping pooled samples or individual samples. One way to reduce the number of genotypings required is to

use pooled samples. A major obstacle in using pooled samples is in terms of accuracy and reproducibility for determining accurate DNA concentrations in setting up the pools. Genotyping individual samples provides higher sensitivity, reproducibility and accuracy and; is the preferred method used in the present invention. Preferably, each individual is genotyped separately and simple gene counting is applied to determine the frequency of an allele of a biallelic marker or of a genotype in a given population.

3) Determining the frequency of a haplotype in case-control populations

The gametic phase of haplotypes is usually unknown when diploid individuals are heterozygous at more than one locus. Different strategies for inferring haplotypes is used to partially 10 overcome this difficulty (see Excoffier L. and Slatkin M., Mol. Biol. Evol., 12(5): 921-927, 1995). One possibility is that the multiple-site heterozygous diploids can be eliminated from the analysis, keeping only the homozygotes and the single-site heterozygote individuals, but this approach might lead to a possible bias in the sample composition and the underestimation of low-frequency haplotypes. Another possibility is that single chromosomes can be studied independently, for example, by asymmetric PCR amplification (see Newton et al., Nucleic Acids Res., 17:2503-2516, 1989; Wu et al., Proc. Natl. Acad. Sci. USA, 86:2757, 1989) or by isolation of single chromosome by limit dilution followed by PCR amplification (see Ruano et al., Proc. Natl. Acad. Sci. USA, 87:6296-6300, 1990). Further, multiple haplotypes can sometimes be inferred using genealogical information in families (Perlin et al., Am. J. Hum. Genet., 55:777-787, 1994). A sample is haplotyped for 20 sufficiently close biallelic markers by double PCR amplification of specific alleles (Sarkar, G. and Sommer S.S., Biotechniques, 1991). These approaches are not entirely satisfying either because of their technical complexity, the additional cost they entail, their lack of generalization at a large scale, or the possible biases they introduce. To overcome these difficulties, an algorithm based on Hardy-Weinberg equilibrium (random mating) to infer the phase of PCR-amplified DNA genotypes 25 introduced by Clark A.G. (Mol. Biol. Evol., 7:111-122, 1990) is used. Briefly, the principle is to start filling a preliminary list of haplotypes present in the sample by examining unambiguous individuals, that is, the complete homozygotes and the single-site heterozygotes. Then other individuals in the same sample are screened for the possible occurrence of previously recognized haplotypes. For each positive identification, the complementary haplotype is added to the list of recognized haplotypes, 30 until the phase information for all individuals is either resolved or identified as unresolved. This method assigns a single haplotype to each multiheterozygous individual, whereas several haplotypes are possible when there are more than one heterozygous site. Any other method known in the art to determine the frequency of a haplotype in a population is used. Preferably, an expectationmaximization (EM) algorithm (Dempster et al., J. R. Stat. Soc., 39B:1-38, 1977) leading to maximum-35 likelihood estimates of haplotype frequencies under the assumption of Hardy-Weinberg proportions is used (see Excoffier L. and Slatkin M., Mol. Biol. Evol., 12(5): 921-927, 1995). The EM algorithm is used to estimate haplotype frequencies in the case when only genotype data from unrelated individuals are available. The EM algorithm is a generalized iterative maximum-likelihood approach to estimation that is useful when data are ambiguous and/or incomplete. The EM algorithm is used to resolve heterozygotes into haplotypes. Haplotype estimations are further described below under the heading "Statistical methods".

4) Genetic Analysis based on Linkage Disequilibrium

Linkage disequilibrium is the non-random association of alleles at two or more loci and represents a powerful tool for genetic mapping of complex traits (see Jorde L.B., Am. J. Hum. Genet., 56:11-14, 1995). Biallelic markers, because they are densely spaced in the human genome and can be genotyped in large numbers, are particularly useful in genetic analysis based on linkage disequilibrium.

When a disease mutation is first introduced into a population (by a new mutation or the immigration of a mutation carrier), it necessarily resides on a single chromosome and thus on a single "background" or "ancestral" haplotype of linked markers. Consequently, there is complete disequilibrium between these markers and the disease mutation: one finds the disease mutation only in the presence of a specific set of marker alleles. Through subsequent generations recombinations occur between the disease mutation and these marker polymorphisms, and the disequilibrium gradually dissipates. The pace of this dissipation is a function of the recombination frequency, so the markers closest to the disease gene will manifest higher levels of disequilibrium than those that are further away. When not broken up by recombination, "ancestral" haplotypes and linkage disequilibrium between marker alleles at different loci can be tracked not only through pedigrees but also through populations.

The pattern or curve of disequilibrium between disease and marker loci will exhibit a single maximum that occurs at the disease locus. Consequently, the amount of linkage disequilibrium between a disease allele and closely linked genetic markers may yield valuable information regarding the location of the disease gene. For fine-scale mapping of a disease locus, it is useful to have some knowledge of the patterns of linkage disequilibrium that exist between markers in the studied region. As mentioned above the mapping resolution achieved through the analysis of linkage disequilibrium is much higher than that of linkage studies. The high density of biallelic markers combined with linkage disequilibrium analysis provide powerful tools for fine-scale mapping. Different methods to calculate linkage disequilibrium are described below under the heading "Statistical Methods". Moreover, association studies as a method of mapping genetic traits rely on the phenomenon of linkage disequilibrium.

3) Association studies

As mentioned above, the occurrence of pairs of specific alleles at different loci on the same chromosome is not random, and the deviation from random is called linkage disequilibrium. If a

specific allele in a given gene is directly involved in causing a particular trait, its frequency will be statistically increased in an affected (trait positive) population when compared to the frequency in a trait negative population or in a random control population. As a consequence of the existence of linkage disequilibrium, the frequency of all other alleles present in the haplotype carrying the traitcausing allele will also be increased in trait positive individuals compared to trait negative individuals or random controls. Therefore, association between the trait and any allele (specifically a biallelic marker allele) in linkage disequilibrium with the trait-causing allele will suffice to suggest the presence of a trait-related gene in that particular allele's region. Association studies focus on population frequencies. Case-control populations can be genotyped for biallelic markers to identify associations that narrowly locate a trait causing allele. Moreover, any marker in linkage disequilibrium with one given marker associated with a trait will be associated with the trait. Linkage disequilibrium allows the relative frequencies in case-control populations of a limited number of genetic polymorphisms (specifically biallelic markers) to be analyzed as an alternative to screening all possible functional polymorphisms in order to find trait-causing alleles. Association studies compare the frequency of marker alleles in unrelated case-control populations, and represent powerful tools for the dissection of complex traits.

Association analysis

10

The general strategy to perform association studies using biallelic markers derived from a candidate gene is to scan two groups of individuals (case-control populations) in order to measure and statistically compare the allele frequencies of the biallelic markers of the present invention in both groups.

If a statistically significant association with a trait is identified for at least one or more of the analyzed biallelic markers, one can assume that: either the associated allele is directly responsible for causing the trait (the associated allele is the trait causing allele), or more likely the associated allele is in linkage disequilibrium with the trait causing allele. The specific characteristics of the associated allele with respect to the candidate gene function usually gives further insight into the relationship between the associated allele and the trait (causal or in linkage disequilibrium). If the evidence indicates that the associated allele within the candidate gene is most probably not the trait causing allele but is in linkage disequilibrium with the real trait causing allele, then the trait causing allele can be found by sequencing the vicinity of the associated marker.

Association studies are usually run in two successive steps. In a first phase, the frequencies of a reduced number of biallelic markers from one or several candidate genes are determined in the trait positive and trait negative populations. In a second phase of the analysis, the identity of the candidate gene and the position of the genetic loci responsible for the given trait is further refined using a higher density of markers from the relevant gene However, if the candidate gene under study is relatively

small in length, as it is the case for many of the candidate genes analyzed included in the present invention, a single phase is sufficient to establish significant associations.

PCT/IB98/02133

Haplotype analysis

As described above, when a chromosome carrying a disease allele first appears in a population as a result of either mutation or migration, the mutant allele necessarily resides on a chromosome having a unique set of linked markers: the ancestral haplotype. This haplotype can be tracked through populations and its statistical association with a given trait can be analyzed. The statistical power of association studies is increased by complementing single point (allelic) association studies with multi-point association studies also called haplotype studies. Thus, a 10 haplotype association study allows one to define the frequency and the type of the ancestral carrier haplotype. A haplotype analysis is important in that it increases the statistical significance of an analysis involving individual markers. Indeed, by performing an association study with a set of biallelic markers, it increases the value of the results obtained through the study, allowing false positive and/or negative data that may result from the single marker studies to be eliminated.

In a first stage of a haplotype frequency analysis, the frequency of the possible haplotypes based on various combinations of the identified biallelic markers of the invention is determined. The haplotype frequency is then compared for distinct populations of trait positive and control individuals. The number of trait positive individuals which should be subjected to this analysis to obtain statistically significant results usually ranges between 30 and 300, with a preferred number of 20 individuals ranging between 50 and 150. The same considerations apply to the number of random control or unaffected individuals used in the study. The results of this first analysis provide haplotype frequencies in case-control populations, the relative risk for an individual carrying a given haplotype of being affected with the given trait under study and the estimated p value for each evaluated haplotype.

25 Interaction Analysis

15

The biallelic markers of the present invention may also be used to identify patterns of biallelic markers associated with detectable traits resulting from polygenic interactions. The analysis of genetic interaction between alleles at unlinked loci requires individual genotyping using the techniques described herein. The analysis of allelic interaction among a selected set of biallelic markers with appropriate level of statistical significance can be considered as a haplotype analysis, similar to those described in further details within the present invention. Preferably, genotyping typing is performed using the microsequencing technique.

Methods to test for association between a trait and a biallelic marker allele or a haplotype of biallelic marker alleles are described below.

XI.D. Statistical methods

In general, any method known in the art to test whether a trait and a genotype show a statistically significant correlation is used.

Methods to estimate haplotype frequencies in a population

As described above, when genotypes are scored, it is often not possible to distinguish heterozygotes so that haplotype frequencies cannot be easily inferred. When the gametic phase is not 5 known, haplotype frequencies can be estimated from the multilocus genotypic data. Any method known to person skilled in the art can be used to estimate haplotype frequencies (see Lange K., Mathematical and Statistical Methods for Genetic Analysis, Springer, New York, 1997; Weir, B.S., Genetic data Analysis II: Methods for Discrete population genetic Data, Sinauer Assoc., Inc., Sunderland, MA, USA, 1996) Preferably, maximum-likelihood haplotype frequencies are computed 10 using an Expectation- Maximization (EM) algorithm (see Dempster et al., J. R. Stat. Soc., 39B:1-38, 1977; Excoffier L. and Slatkin M., Mol. Biol. Evol., 12(5): 921-927, 1995). This procedure is an iterative process aiming at obtaining maximum-likelihood estimates of haplotype frequencies from multi-locus genotype data when the gametic phase is unknown. Haplotype estimations are usually performed by applying the EM algorithm using for example the EM-HAPLO program (Hawley M.E. et al., Am. J. Phys. Anthropol., 18:104, 1994) or the Arlequin program (Schneider et al., Arlequin: a software for population genetics data analysis, University of Geneva, 1997). The EM algorithm is a generalized iterative maximum likelihood approach to estimation and is briefly described below.

In the following part of this text, phenotypes will refer to multi-locus genotypes with unknown phase. Genotypes will refer to known-phase multi-locus genotypes.

Suppose a sample of N unrelated individuals typed for K markers. The data observed are the unknown-phase K-locus phenotypes that can categorized in F different phenotypes. Suppose that we have H underlying possible haplotypes (in case of K biallelic markers, $H=2^K$).

For phenotype j, suppose that cj genotypes are possible. We thus have the following equation

25
$$P_j = \sum_{i=1}^{c_j} pr(genotype_i) = \sum_{i=1}^{c_j} pr(h_k, h_l)$$
 Equation 1

where Pj is the probability of the phenotype j, hk and hl are the two haplotypes constituent the genotype i. Under the Hardy-Weinberg equilibrium, pr(hk,hl) becomes:

$$pr(h_k, h_l) = pr(h_k)^2$$
 if $h_k = h_l$, $pr(h_k, h_l) = 2pr(h_k) \cdot pr(h_l)$ if $h_k \neq h_l$. Equation 2

The successive steps of the E-M algorithm can be described as follows:

Starting with initial values of the of haplotypes frequencies, noted, $p_1^{(0)}$, $p_2^{(0)}$,...., $p_T^{(0)}$. these initial values serve to estimate the genotype frequencies (Expectation step) and then estimate another set of haplotype frequencies (Maximization step): $p_1^{(1)}$, $p_2^{(1)}$,...., $p_T^{(1)}$. these two steps are iterated until change in the sets of haplotypes frequency are very small.

A stop criterion can be that the maximum difference between haplotype frequencies between two iterations is less than 10⁻⁷. These values can be adjusted according to the desired precision of estimations.

In detail, at a given iteration s, the Expectation step consists in calculating the genotypes frequencies by the following equation:

$$pr(genotype_{i})^{(s)} = pr(phenotype_{j}).pr(genotype_{i}|phenotype_{j})^{(s)}$$

$$= \frac{n_{j}}{N} \cdot \frac{pr(h_{k}, h_{l})^{(s)}}{P_{j}^{(s)}}$$
Equation 3

where genotype i occurs in phenotype j, and where hk and hl constitute genotype i. Each probability are derived according to equations 1 and 2 above.

Then the Maximization step simply estimates another set of haplotype frequencies given the genotypes frequencies. This approach is also known as gene-counting method (Smith, Ann. Hum. Genet., 21:254-276, 1957).

$$p_t^{(s+1)} = \frac{1}{2} \sum_{i=1}^{F} \sum_{i=1}^{c_i} \delta_{it} . pr(genotype_i)^{(s)}$$
 Equation 4

where δ_{it} is an indicator variable which count the number of time haplotype t in genotype i. It takes the values of 0, 1 or 2.

To ensure that the estimation finally obtained are the maximum-likelihood estimations several values of departures are required. The estimations obtained are compared and if they differ the estimations leading to the best likelihood are kept. The term "haplotype determination method" is used to refer to all methods for determinin haplotypes known in the art including expectation-maximization algorithms.

20 Methods to calculate linkage disequilibrium between markers

A number of methods can be used to calculate linkage disequilibrium between any two genetic positions, in practice, linkage disequilibrium is measured by applying a statistical association test to haplotype data taken from a population.

Linkage disequilibrium between any pair of biallelic markers comprising at least one of the biallelic markers of the present invention (M_{i1}, M_{j}) can be calculated for every allele combination $(M_{i1}, M_{j1}; M_{i1}, M_{j2}; M_{i2}, M_{j1})$ and M_{i2}, M_{j2} , according to the Piazza formula:

$$\Delta \mathbf{M}_{ik}$$
, \mathbf{M}_{ji} = $\sqrt{\theta}4 - \sqrt{(\theta}4 + \theta}3)(\theta}4 + \theta}2)$, where :

 $\theta 4 = -- =$ frequency of genotypes not having allele k at M_i and not having allele l at M_j

 θ 3= - + = frequency of genotypes not having allele k at M_i and having allele l at M_j

30 θ_{2} = + - = frequency of genotypes having allele k at M_i and not having allele 1 at M_j

Linkage disequilibrium (LD) between pairs of biallelic markers (Mi, Mj) can also be calculated for every allele combination (M_{i1} , M_{j1} ; M_{i1} , M_{j2} ; M_{i2} , M_{j1} and M_{i2} , M_{j2}), according to the maximum-likelihood estimate (MLE) for delta (the composite linkage disequilibrium coefficient), as described by Weir (B.S. Weir, *Genetic Data Analysis*, Sinauer Ass. Eds, 1996). This formula allows linkage disequilibrium between alleles to be estimated when only genotype, and not haplotype, data are available. This LD composite test makes no assumption for random mating in the sampled population, and thus seems to be more appropriate than other LD tests for genotypic data.

Another means of calculating the linkage disequilibrium between markers is as follows. For a couple of biallelic markers, Mi (a/b_i) and Mj (a/b_j) , fitting the Hardy-Weinberg equilibrium, one can estimate the four possible haplotype frequencies in a given population according to the approach described above.

The estimation of gametic disequilibrium between ai and aj is simply:

$$D_{aiai} = pr(haplotype(a_i, a_j)) - pr(a_i).pr(a_j).$$

Where pr(ai) is the probability of allele ai and aj is the probability of allele aj, and where $pr(haplotype\ (ai,\ aj))$ is estimated as in Equation 3 above.

For a couple of biallelic marker only one measure of disequilibrium is necessary to describe the association between Mi and Mj.

Then a normalized value of the above is calculated as follows:

The skilled person will readily appreciate that other LD calculation methods can be used without undue experimentation.

Linkage disequilibrium among a set of biallelic markers having an adequate heterozygosity rate can be determined by genotyping between 50 and 1000 unrelated individuals, preferably between 25 75 and 200, more preferably around 100.

Testing for association

Methods for determining the statistical significance of a correlation between a phenotype and a genotype, in this case an allele at a biallelic marker or a haplotype made up of such alleles, is determined by any statistical test known in the art and with any accepted threshold of statistical significance being required. The application of particular methods and thresholds of significance are well with in the skill of the ordinary practitioner of the art.

Testing for association is performed by determining the frequency of a biallelic marker allele in case and control populations and comparing these frequencies with a statistical test to determine if their is a statistically significant difference in frequency which would indicate a correlation between the trait and the biallelic marker allele under study. Similarly, a haplotype analysis is performed by estimating the frequencies of all possible haplotypes for a given set of biallelic markers in case and

control populations, and comparing these frequencies with a statistical test to determine if their is a statistically significant correlation between the haplotype and the phenotype (trait) under study. Any statistical tool useful to test for a statistically significant association between a genotype and a phenotype is used. Preferably the statistical test employed is a chi square test with one degree of freedom. A P-value is calculated (the P-value is the probability that a statistic as large or larger than the observed one would occur by chance).

Statistical significance

15

In preferred embodiments, significance for diagnosis purposes, either as a positive basis for further diagnostic tests or as a preliminary starting point for early preventive therapy, the p value related to a biallelic marker association is preferably about 1 x 10-2 or less, more preferably about 1 x 10-4 or less, for a single biallelic marker analysis and about 1 x 10-3 or less, still more preferably 1 x 10-6 or less and most preferably of about 1 x 10-8 or less, for a haplotype analysis involving several markers. These values are believed to be applicable to any association studies involving single or multiple marker combinations.

The skilled person can use the range of values set forth above as a starting point in order to carry out association studies with biallelic markers of the present invention. In doing so, significant associations between the biallelic markers of the present invention and cancer and prostate cancer can be revealed and used for diagnosis and drug screening purposes.

Using the method described above and evaluating the associations for single marker alleles or for haplotypes permits an estimation of the risk a corresponding carrier has to develop a given trait, and particularly in the context of the present invention, a disease, preferably cancer, more preferably prostate cancer. Significance thresholds of relative risks are to be adapted to the reference sample population used.

In this regard, among all the possible marker combinations or haplotypes which are evaluated to determine the significance of their association with a given trait, for example a form of cancer or prostate cancer, a response to treatment with anti-cancer or anti-prostate cancer agents or side effects related to treatment with anti-cancer or anti-prostate cancer agents, it is believed that those displaying a coefficient of relative risk above 1, preferably about 5 or more, preferably of about 7 or more are indicative of a "significant risk" for the individuals carrying the identified haplotype to develop the given trait. It is difficult to evaluate accurately quantified boundaries for the so-called "significant risk". Indeed, and as it has been demonstrated previously, several traits observed in a given population are multifactorial in that they are not only the result of a single genetic predisposition but also of other factors such as environmental factors or the presence of further, apparently unrelated, haplotype associations. Thus, the evaluation of a significant risk must take these parameters into consideration in order to, in a certain manner, weigh the potential importance of external parameters in the developm nt of a given trait. Without wishing to be bound to any invariable model or theory

PCT/IB98/02133 WO 99/32644 131

based on the above statistical analyses, the inventors believe that a "significant risk" to develop a given trait is evaluated differently depending on the trait under consideration.

It will of course be understood by practitioners skilled in the treatment or diagnosis of cancer and prostate cancer that the present invention does not intend to provide an absolute identification of 5 individuals who could be at risk of developing a particular form of cancer or who will or will not respond or exhibit side effects to treatment with anti-cancer or anti-prostate cancer agents but rather to indicate a certain degree or likelihood of developing a disease or of observing in a given individual a response or a side effect to treatment with a particular agent or set of agents.

However, this information is extremely valuable as it can, in certain circumstances, be used to 10 initiate preventive treatments or to allow an individual carrying a significant haplotype to foresee warning signs such as minor symptoms. In the case of cancer, the knowledge of a potential predisposition, even if this predisposition is not absolute, might contribute in a very significant manner to treatment, or allow for suggestions in changes in diet or the reduction of risky behaviors, e.g. smoking. Similarly, a diagnosed predisposition to a potential side effect could immediately direct 15 the physician toward a treatment, for which such side effects have not been observed during clinical trials.

Phenotypic randomization

In order to confirm the statistical significance of the first stage haplotype analysis described above, it might be suitable to perform further analyses in which genotyping data from case-control 20 individuals are pooled and randomized with respect to the trait phenotype. Each individual genotyping data is randomly allocated to two groups which contain the same number of individuals as the case-control populations used to compile the data obtained in the first stage. A second stage haplotype analysis is preferably run on these artificial groups, preferably for the markers included in the haplotype of the first stage analysis showing the highest relative risk coefficient. This experiment is reiterated between 50 and 200 times, preferably between 75 and 125 times. The repeated iterations allow the determination of the percentage of obtained haplotypes with a significant p-value level below about 1x10-3.

Example 24

Detailed Association Studies

The initial association studies between the 8p23 locus and prostate cancer described in Section I.D. were repeated at a higher level of sophistication.

Collection of DNA samples from affected and non-affected individuals

Prostate cancer patients were recruited according to clinical inclusion criteria based on pathological or radical prostatectomy records as described above in Section I. However, the pool of 35 individuals suffering from prostate cancer described in Section I was augmented from the original 185 individuals to a range of between 275 and 491 individuals depending on the marker tested. Similarly,

the control pool of non-diseased individuals described in Section I was augmented from the original 104 individuals to a range of between 130 and 313 individuals depending on the marker tested.

Genotyping Affected and Control Individuals

As for Section I.D., allelic frequencies of the biallelic markers in each population were determined by performing microsequencing reactions on amplified fragments obtained by genomic PCR performed on the DNA samples from each individual as described in Example 5.

Association Studies

Association results were obtained using markers spanning a 650 kb region of the 8p28 locus around PG1 both using single point analysis and haplotyping studies. See Figure 16. As compared with the earlier representation of the initial association results for this region shown in Figure 2, Figure 16 is to scale, since the entire region has now been sequenced. In addition, more markers were generated around the association peak in the area of PG1; each of which has been tested in single point analysis (hence the density of data within this subregion). The haplotyping curve in Figure 16 represents, for each marker considered, the maximum p-value for haplotypes obtained using this marker and any number from all markers harbored by the same BAC and being in Hardy Weindeberg Disequilibrium with said marker.

The data presented in Figure 16 shows a strong association between this specific region within 8p23 locus, especially in the area that has been identified as being the PG1 gene, and prostate cancer. The maximum p-value in single point analysis, for the PG1 sub-region, is 3.10⁻³, while outside of the PG1 subregion, most of the p-values obtained for single point associations are less significant than 1.10⁻¹. The maximum p-value obtained for haplotyping studies is the one obtained for a marker inside PG1's BAC, and equals 3.10⁻⁶.

Figure 17 is a graph showing an enlarged view of the single point association results within a 160 kb region comprising the PG1 gene. Markers involved in this enlargement were all located on BAC B0463F01 (see Figure 16), except marker 4-14, which lies in very close proximity, on BAC B0189E08. Figure 17 shows all of the markers which made up the maximum haplotype shown in Figure 16. Some of these markers were later revealed to lie within the promoter, exonic or intronic regions of the PG1 gene. The markers outside the gene were all informative biallelic markers with a least frequent allele present at a frequency of more than 20%, while markers within the gene were a mix of such informative markers and markers whose least frequent allele's frequency is less than 20%. These data confirm and narrow the previous peak of association values seen in Figure 16, to a 40 kb harboring the PG1 gene. Significant associations are obtained for markers starting at the promoter site with marker No. 99-1485, and ending at the 3' UTR site with marker No. 5-66.

Figure 18A is a graph showing an enlarged view of the single point association results of 40 kb within the PG1 gene. These data confirm that seven markers within the PG1 gene have one allele associated with prostate cancer, with p-values all similar and more significant than 1.10⁻², specifically

markers 99-622; 4-77; 4-71; 4-73; 99-598; 99-576; 4-66. Figure 18B is a table listing the location of markers within PG1 gene, the two possible alleles at each site. For each marker, the disease-associated allele is indicated first; its frequencies in cases and controls as well as the difference between both are shown; the odd-ratio and the p-value of each individual marker association are also shown.

The data in Figures 17, 18A, and 18B demonstrate that the markers in the PG1 gene have an association with prostate cancer that is valid, and exhibits similar significance values, regardless whether the considered cases are sporadic or familial cases. Therefore, some PG1 alleles must be general risk factors for any type of prostate cancer, whether familial or sporadic. The fact that several p-values for associated alleles are around 1.10⁻² suggests that all these markers are in linkage disequilibrium to one another, and can all be used individually to assess PG1 associated prostate cancer susceptibility risk. The prostate cancer associated alleles of the 7 markers discussed above, all exhibit an odd-ratio of about 1.5, which means for each of them that an individual carrying such allele has 1.5 more chances to be susceptible to prostate cancer than not.

In order to confirm the significance of the association results found for markers on the BAC harboring PG1, we a novel statistical method was performed as described in provisional patent application serial no. 60/107,986, filed November 10, 1998.

Haplotype analysis

The results of a haplotype analysis study using 4 markers (marker Nos. 4-14, 99-217, 4-66 and 99-221)) within the 160 kb region shown in Figure 17 are shown in Figure 19A. These 4 markers have each been shown to be strongly associated with prostate cancer, i.e. with p-values more significant than 1.10⁻³ on approximately 150 cases and 130 controls. All haplotypes using 2, 3, or 4 markers among the 4 above cited were analyzed using 491 case patients and 317 control individuals. Figure 19A shows the most significant haplotypes obtained, as well as the individual odd-ratios for each. Haplotype 11 is the most significant (p-value of ca. 3.10⁻⁶), and is related to haplotype 5, shown in Figure 4 in that three of the four marker alleles (4-14 C, 99-217 T and 99-221 A) are common to both haplotypes, and both cover a similar region. Differences in p-values are explained both by the addition of markers and of more case or control individuals. Haplotype 11 has an highly informative odd-ratio (of above 3); it is present in 3% of the controls and almost 10% of the cases.

Figure 19B is a table showing the segmented haplotyping results according to the age of the subjects, and whether the prostate cancer cases were sporadic or familial, using the same markers 4 markers and the same individuals as were used to generate the results in Figure 19A. Figure 19B shows equivalent results for all segments of the population analyzed, demonstrating that the PG1 associated alleles are general risk factors for prostate cancer, regardless of the age of onset of the disease.

30

35

5

10

PCT/IB98/02133

134

The haplotyping results and odd ratios for all of the combinations of the 7 markers (99-622; 4-77; 4-71; 4-7; 99-598; 99-576; and 4-66) within PG1 gene that were shown in Figure 18 to have p-values more significant than 1 x 10⁻² were computed. A portion of these data are shown in Figure 20. All of the 2-, 3-, 4-, 5-, 6- and 7-marker haplotypes were tested. Figure 20 identifies for each x-marker haplotype category, the most significant haplotype. Among all these, the most significant haplotype is the two-marker haplotype 1, which shows a p-value of approximately 6.10⁻⁵, with an odd ratio of 2. The frequency of haplotype 1 among the control individuals is 15%, while it is 26% among the case patients. It is worth noting that these frequencies are very similar for all haplotypes presented on Figure 20. It will thus be sufficient to test this two marker haplotype for prognosis/diagnosis on risk patients, as opposed to having a more complex test of a haplotype comprising 3 or more makers.

Finally, Figure 21 is a graph showing the distribution of statistical significance, as measured by Chi-square values, for each series of possible x-marker haplotypes, (x =2, 3 or 4) using all of the 19 markers found in PG1 gene. These data confirm that testing 2-marker haplotypes within PG1 is sufficient because the testing 3- or 4-marker haplotypes does not increase the statistical relevance of the analysis.

Example 25

Attributable Risk

Attributable risk describes the proportion of individuals in a population exhibiting a phenotype due to exposure to a particular factor. For further discussion of attributable risk values, see 20 Holland, Bart K., Probability without Equations – Concepts for Clinicians; The Johns Hopkins University Press, pp. 88-90. In the present case the phenotype examined was prostate cancer, and the exposure was either one single allele of an individual PG1-related marker, or a haplotype thereof in an individual's genome.

The formula used for calculating attributable risk values in the present study was the following:

25 $AR = P_E (RR-1) / [P_E (RR-1)+1], where:$

AR was the attributable risk of allele or haplotype;

P_E was the frequency of exposure to allele or haplotype within the population at large, in the present study a random male Caucasian population; and

RR was the relative risk, in the present study relative risk is approximated with the odd-ratio,

because of the relatively low incidence of prostate cancer in populations at large (values for the odd ratios are found in Figures 18B and 20).

In this case, P_E was estimated using a dominant transmission model for prostate cancer:

 $P_E = (N_{AA} + N_{AB}) / N$, where:

N_{AA} was the number of homozygous individuals harboring the disease associated allele or haplotype within a given random population, and N_{AB} was the number of heterozygous individuals is said random population. N_{AA} and N_{AB} were calculated using the allele frequencies in the random

population as indicated in Figures 18B and 20, and N was the number of individuals in total random population.

We calculated the attributable risks of disease-associated alleles for markers within PG1 gene and presented these results in Figure 18B. In Figure 20, the attributable risk for the two-marker haplotypes present in the figure as shown as well. These data demonstrate that disease-associated alleles of PG1 are present in approximately 20% of prostate cancer patients in the Caucasian population at large, and therefor represent prognostic tools of significant value.

SEQUENCE LISTING FREE TEXT

10 The following free text appears in the accompanying Sequence Listing:

identification method Proscan

potential start codon

exon1

Tyr phos

15 upstream amplification primer

polymorphic fragment

polymorphic base

downstream amplification primer

complement

20 upstream amplification primer 99-217-PU, extracted from SEQ ID1 34216 34234

Klein, Kanehisa and DeLisi identification method, potential helix

Eisenberg, Schwarz, Komarony, Wall identification method, potential helix

Prosite match

potential Tyrosine kinase site, Prosite match

25 potential caseine kinase II site, Prosite match

potential Leucine zipper site, Prosite match

potential site, Prosite match

potential protein kinase C, Prosite match

potential cAMP and cGMP dependant protein kinase site, Prosite match

30 primer oligonucleotide

box2 from SEQID4, present in AF003136, P33333, P26647, U89336, U56417, AB005623

box2 from Z72511

box3 from SEQID4, present in AF003136

35 potential microsequencing oligo

complenent potential microsequencing oligo

136

polymorphic fragment 4-77, extracted from SEQ ID1 12057 12103 polymorphic fragment 99-123, variant version of SEQ ID21 base A; G in SEQ ID22

downstream amplification primer 99-217-RP, extracted from SEQ ID1 34625 34645

5 complement

polymorphic base C in PG1 (13680) SEQ ID1

stop codon

potential

amplification oligonucleotide

10 sequencing oligonucleotide

Box II

Box III

Box I

upstream amplification primer for SEQ 188, SEQ 265, SEQ 189, SEQ 266

downstream amplification primer for SEQ 185, SEQ 262, SEQ 186, SEQ 263, SEQ 187, SEQ

264

microsequencing oligo for 4-20-149.mis1

Although this invention has been described in terms of certain preferred embodiments, other embodiments which will be apparent to those of ordinary skill in the art in view of the disclosure herein are also within the scope of this invention. Accordingly, the scope of the invention is intended to be defined only by reference to the appended claims.

Table 1

marker.	polymorphism	most frequent	less frequent	cases		controls	
				p *	q**	p*	q**
99-123	СЛ	С	T	0,65	0,35	0,7	0,3
4-26	A/G	A	G	0,61	0,39	0,55	0,45
4-14	С/Т	С	T	0,65	0,35	0,59	0,41
4-77	C/G	С	G	0,67	0,33	0,76	0,24
99-217	С/Т	C	T	0,69	0,31	0,77	0,23
4-67	СЛ	С	T	0,74	0,26	0,84	0,16
99-213	A/G	Α '	G	0,55	0,45	0,62	0,38
99-221	C/A	[c	Α !	0,43	0,57	0,43	0,57
99-135	A/G	A	G	0,75	0,25	0,7	0,3

^{*:} frequency of most frequent base within each sub-population
**: frequency of least frequent base within each sub-population (p+q=1) standard deviations ~0,023 to 0,031 for controls standard deviations ~0,018 to 0,021 for cases

WHAT IS CLAIMED IS:

- 1. A recombinant, purified or isolated polynucleotide comprising a mammalian PG1 gene, cDNA, complement thereof, or fragment thereof having at least 10 nucleotides in length.
- 2. The polynucleotide according to claim 1, wherein said mammalian PG1 gene or cDNA is human or mouse.
- The polynucleotide according to claim 2, wherein the polynucleotide is selected from SEQ ID NOs: 3, 69, 112-124, 179, and 182-184.
 - 4. A polynucleotide selected from SEQ ID NOs: 185-578.
- 5. A purified or isolated polypeptide comprising a mammalian PG1 protein, or fragment thereof having at least 8 amino acids in length.
 - 6. The polypeptide according to claim 5, wherein said mammalian PG1 protein is human or mouse.
- The polypeptide according to claim 6, wherein said polypeptide is selected from SEQ ID NOs: 4, 5, 70, 74, and 125-136.
 - 8. The polypeptide according to claim 5, wherein said polypeptide consists of said mammalian PG1 protein, or fragment thereof having at least 8 amino acids in length.
 - 9. A polynucleotide comprising a nucleic acid sequence encoding a polypeptide according to claim 8.
- 10. An antibody composition capable of selectively binding to an epitope-containing 30 fragment of a polypeptide according to claim 8, wherein said antibody is either polyclonal or monoclonal.
 - 11. A vector comprising a polynucleotide according to any one of claims 1, 4, and 9.
- 35 12. A host cell comprising a polynucleotide according to claim 11.

- 13 A nonhuman host animal or mammal comprising a vector according to claim 11.
- 14. A mammalian host cell comprising a PG1 gene disrupted by homologous recombination with a knock out vector.

- 15. A nonhuman host mammal comprising a PG1 gene disrupted by homologous recombination with a knock out vector.
- 16. A polynucleotide according to any one of claims 1, 4, and 9, further comprising a 10 label.
 - 17. A polynucleotide according to any one of claims 1, 4, and 9, attached to a solid support.
- 15 18. A random or addressable array of polynucleotides comprising at least one polynucleotide according to any one of claims 1, 4, and 9.
 - 19. A method of determining whether an individual is at risk of developing cancer or prostate cancer, or whether said individual suffers from cancer or prostate cancer as a result of a mutation in the PG1 gene comprising:

obtaining a nucleic acid sample from said individual; and

determining whether the nucleotides present at one or more PG1-related biallelic marker are indicative of a risk of developing cancer or prostate cancer or indicative of cancer or prostate cancer resulting from a mutation in the PG1 gene.

25

20. A method of determining whether an individual is at risk of developing cancer or prostate cancer or whether said individual suffers from cancer or prostate cancer as a result of a mutation in the PG1 gene comprising:

obtaining a nucleic acid sample from said individual; and

- determining whether the nucleotides present at one or more PG1-related biallelic marker are indicative of a risk of developing cancer or prostate cancer or indicative of cancer or prostate cancer resulting from a mutation in the PG1 gene.
- 21. A method according to either one of claims 19 and 20, wherein said PG1-related 35 biallelic is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-

77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66.

5 22. A method of obtaining an allele of the PG1 gene which is associated with a detectable phenotype comprising:

obtaining a nucleic acid sample from an individual expressing said detectable phenotype; contacting said nucleic acid sample with an agent capable of specifically detecting a nucleic acid encoding the PG1 protein; and

isolating said nucleic acid encoding the PG1 protein.

23. A method of obtaining an allele of the PG1 gene which is associated with a detectable phenotype comprising:

obtaining a nucleic acid sample from an individual expressing said detectable phenotype;

contacting said nucleic acid sample with an agent capable of specifically detecting a sequence within the 8p23 region of the human genome;

identifying a nucleic acid encoding the PG1 protein in said nucleic acid sample; and isolating said nucleic acid encoding the PG1 protein.

- 20 24. A method of categorizing the risk of prostate cancer in an individual comprising the step of assaying a sample taken from the individual to determine whether the individual carries an allelic variant of PG1 associated with an increased risk of prostate cancer.
 - 25. The method of Claim 24 wherein said sample is a nucleic acid sample.

- 26. The method of Claim 24 wherein said sample is a protein sample.
- 27. The method of Claim 26, further comprising determining whether the PG1 protein in said sample binds an antibody that binds specifically to a PG1 isoform associated with prostate 30 cancer.
 - 28. A method of genotyping comprising determining the identity of a nucleotide at a PG1-related biallelic marker in a biological sample.

29. A method of estimating the frequency of an allele in a population comprising determining the proportional representation of a nucleotide at a PG1-related biallelic marker in a pooled biological sample derived from said population.

- 5 30. A method of detecting an association between a genotype and a phenotype, comprising the steps of:
 - a) genotyping at least one PG1-related biallelic marker in a trait positive population;
 - b) genotyping said PG1-related biallelic marker in a control population; and
- c) determining whether a statistically significant association exists between said genotype and said phenotype.
 - 31. A method of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising:
 - a) genotyping at least one PG1-related biallelic marker;
- b) genotyping a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the genome of each individual in said population; and
 - c) applying an haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency.

32. A method of detecting an association between a haplotype and a phenotype, comprising the steps of:

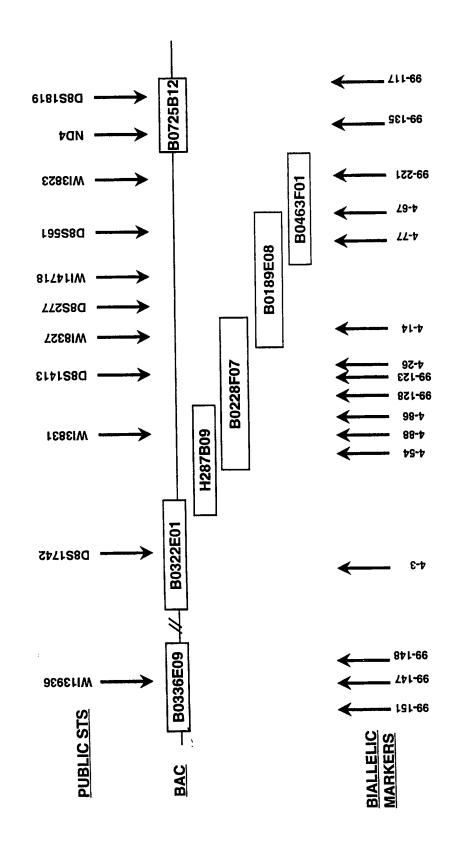
- a) estimating the frequency of at least one haplotype in a trait positive population according to the method of claim 31;
- b) estimating the frequency of said haplotype in a control population according to the method of claim 31; and
 - c) determining whether a statistically significant association exists between said haplotype and said phenotype.
- 30 33. A method according to claim 31, wherein said PG1-related biallelic marker and said second biallelic marker are 4-77/151 and 4-66/145,
 - 34. A method according to claim 32, wherein said haplotype exhibits a p-value of < 1x 10^{-3} in an association with a trait positive population with cancer, or prostate cancer.

- 35. A method according to any one of claims 29 to 31, wherein said PG1-related biallelic is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66.
 - 36. A method according to either one of claims 30 and 32, wherein said control population is a trait negative population or a random population.

- 37. A method according to any one of claims 22, 23, 30, and 32, wherein said phenotype is a disease, cancer or prostate cancer; a response to an anti-cancer agent or an anti-prostate cancer agent; or a side effect to an anti-cancer or anti-prostate cancer agent.
- 15 38. A polynucleotide for use in a hybridization assay for determining the identity of the nucleotide at an PG1-related biallelic marker; for use in a sequencing assay for determining the identity of the nucleotide at an PG1-related biallelic marker; for use in a allele-specific amplification assay for determining the identity of the nucleotide at an PG1-related biallelic marker; or for use in amplifying a segment of nucleotides comprising an PG1-related biallelic marker.

20

39. The polynucleotide according to claim 38, wherein said PG1-related biallelic is a PG1-related biallelic markers positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576, and 4-66.



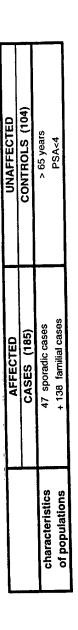
NSDOCID: <WO___9932644A2_I_>

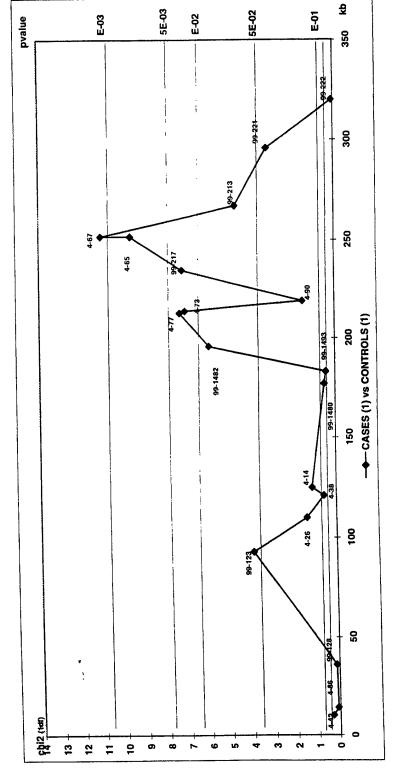
ASSOCIATION STUDIES (FIRST SCREENING)

NON AFFECTED	CONTROLS=76	> 65 years	PSA<4	į
AFFECTED	CASES = 112	35 sporadic cases	+ 77 familial cases	
Population	Sample Size	Population	Characteristics	

pvalue	5E-04	E-03	5E-03	E-02	5E-02	<u> </u>	
						99-117	-
	•			;		99-135	
	194		 			99-221	
		E.	; !	1			
	ZOOM	4-14		+			
					4-26	99-123	controls
						95-128	-delta allelic frequency cases vs controls
	•					488	ilelic freque
						99-148	' -
						96	
chi2	4 61	7 = 0	+ +	+ + 9	· 70 4	8 2 -	•

ASSOCIATION STUDIES (ZOOM)





HAPLOTYPE FREQUENCY ANALYSIS

POPULATIONS

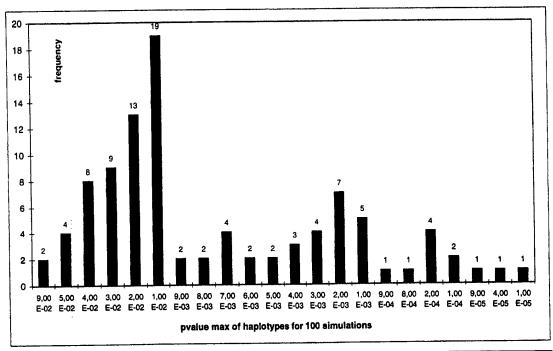
UNAFFECTED
CONTROLS 3 (130)
> 65 years
PSA<4 AFFECTED
CASES 2 (281)
143 sporadic cases
+ 138 familial cases characteristics of populations

		_			_	4/2	4				_			-	,			
		pvalue						9,00E-04 ***	6,00E-05 ****	1,00E-05 *****	9,00E-07 *****	2,00E-05 *****	2,00E-05 *****	4,00E-05 ****	2,00E-04 ****	1,00E-04 ***	3,00E-04 ****	6,00E-04 ***
			;	relative	risk			4.42	6.46	6.78	10,06	5,17	4,78	2,33	2,17	2,32	2,01	2,05
	haplotype	projec	ויבלתפווכוכי		controls	(6)		0.018	0.016	0.019	0,013	0,025	0,027	0,109	0,134	0,112	0,146	0,129
_	hap	trodi	ביל ביל		cases			0.075	0.095	0.116	0.117	0,117	0.117	0,222	0.251	0,226	0,256	0,233
99-135	80725812				2 00F-01		>100kb<	4	∶ ∢	:								
99-221					7 OOF-01		<29kb>	4	∶ ⊲	< ∢	< ∢	∶ ∢	< <					
99-213					9 00E-02	3,100,6	<15kb>	ď	י ני	ט פ	ා ර	J (C	. c	. C	i	Ø	j.	g
4-67	DOMESTICAL	204021		Î	VO 200 9		<17kb>	F	- 1-		- }-				- -		. ⊢	. ⊢
99-217			11453	PG1	20.00	z)-ann-2	<22kb>	۲	- +	- 1-	- 1-	- ⊢	- -		-		- +-	-
4-77			114	,	00000	2,00E-02	/88kh>		5 (5 (5 C	5 (3	ď	5 C	5		
4-14		B0189E08				1,005-01	AELP	Z I SWELV	، د	ပ (، د	د						
4-26		BO1				1,00E-01	1		⋖ •	∢ ⋅	∢							
99-123	33.150	H0287B09				2,00E-01			ပ									
	markers	bacs	on journ	Soluiga	genes	p value		distance between markers (kb)	haplotype 8 >304kb<	haplotype 7 >286kb<	haplotype 6 <186kb>	haplotype 5 <171kb>	haplotype 4 <83kb>	haplotype 3.1 <54kb>	haplotype 3.2 <54kb>	hapiotype 2.2 <39kb>	haplotype 2 <32kb>	haplotype 1.1 <17 kb>

Figure 5

HAPLOTYPE SIMULATIONS (100 ITERATIONS)

							haplotype	frequencies	relativ	pvalue
markers	4-14	4-77	99-217	4-67	99-213	99-221	cases	controls	risk	
haplotype	С	G	Т	Т	G	Α	0,117	0,013	10,06	9,00E-07



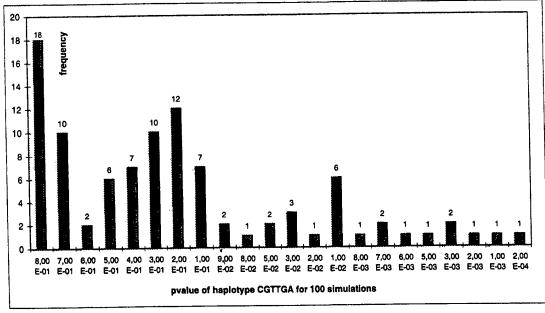


Figure 6A

Microseq. oligos positions	1-23 25-47 (complementary)
Base	C/T N/G C/T C/T C/T A/G
SEQ ID N° Polymorphism position*	22 24 24 24 24 24 24
SEQ ID N°	48 49 52 53 53 56 56
RP sequence	TATTCAGAAAGGAGTGGG TGAGGACTGCTAGGAAAG GACTGTATCCTTTGATGCAC GGAAAGGTACTCATTCATAG GTTTATTTGTGTGAGCTTTG TGAAAGAGTTATTCTCTGG TTATTGCCCCAATGCTGAG TCATTCGTCTGAGCTGAG
SEO ID Nº	39 4 4 4 4 4 4 4 6 4 6 4 6
PU sequence	AAAGCCAGGACTAGAAGG TACAGCCCTGTAAGACAC TCTAACCTCTCATCCAAC TGTTGATTTACAGGCGGC GGTGGGAATTTACTATATG AAGTTCACCTTCTCAAGC ATACTGGCAGCGTGTGCTTC CCCTTTTTCTTATTGCTTCCAAGC TGTGGCAGCGTGTGCTTC TGGAAGTTGTTATTGCCTTCCAAGC TGGAAGTTGTTATTGCCCC
SEQ ID N° (mut)	8 2 8 8 8 8 8 8
SEQ ID N° SEQ ID N	28848888
Marker	99-123 4-26 4-14 4-77 99-217 99-213 99-221
BAC	228 189 228/189 189/463 463 463 463 725

Figure 6B

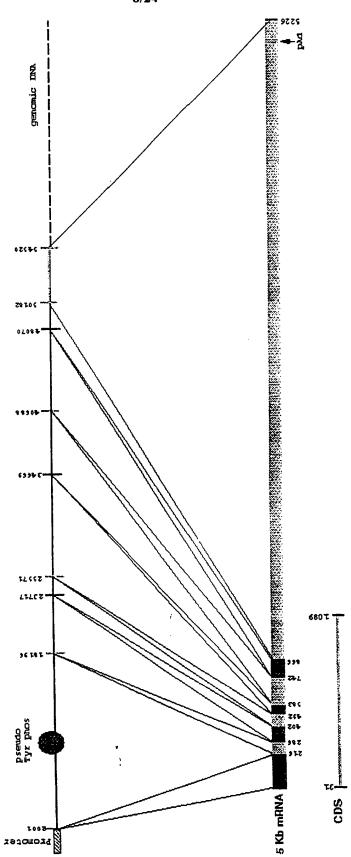
25-47 (complementary) 25-47 (complementary) 25-47 (complementary) Microseq. oligos positions 2 2 2 2 2 3 882 Polymorphism position* 2 2 2 SEO ID N° 66 67 68 ACAAATCTATATAAGGCTGG CTCTTGGTTAAACAGCAGTG TGGCTCTGCATTTCTTCC RP sequence SEO ID Nº 8 2 8 ATCAAATCAGTGAAGTCTGAG ATCGCTGGAACATTCTGG GATTTAAGCTACGCTATTAG PU sequence SEO ID Nº SEO ID Nº (mut 8 2 8 25 28 25 Marker 99-1482 4-73 4-65 189/463 463 463 BAC

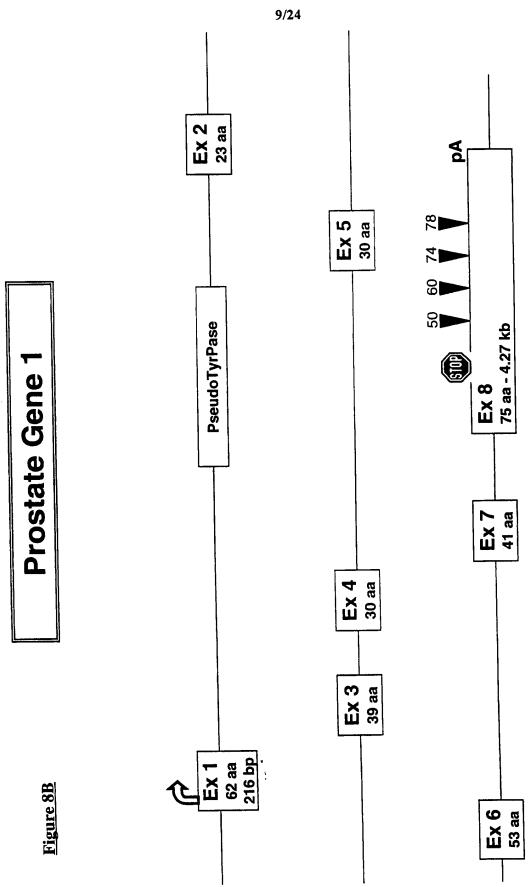
•: positions are given relative to the sequence of the corresponding marker (i.e. SEQ ID N° 21-38 and 57-62)

Figure 7

EXON Phase		START	END	5' SPsite	PHASE	3' SPsite
Ex1 +0		2001	2216			GTGAGC
Ex2	GTTTGTA	18196	18265	TAG	+0	
+1						
Ex3 +0		23717	23832	CAG	+2	GTAACT
Ex4	GTAAGA	25571	25660	CAG	+0	
+0						
Ex5 GTAAC +2	GT .	34669	34759	CAG	+0	
Ex6 +1		40688	40846	TAG	+1	GTAAGT
Ex7 +2		48070	48193	TAG	+2	GTGAGT
Ex8		50182	54523	TAG	+1	
ATG co	odon	2031	2033			
STOP	codon	50405	50407			
POLY	Ad site	54445	54450			

PG1 genome structure and cDNAs





10/24

		box1	box2	box3
PG1	Hs	NHQ 81-83	FPEGTR 160 –165	LDAIYDVTV 211-219
AF003136 (Genbank)	Ce	NHQ 630 -632	FPEGTR 712-717	LDAIYDVTV 762-770
Z72511 (Genbank)	Ce	48 NHR 50	FPEGTD 129-134	VEYIYDITI 204-212
P38226 (Swissprot)	Sc	111 NHQ 113	FPEGTN 223-228	IESLYDITI 271-279
P33333 (Swissprot)	Sc	81 NHQ 83	FPEGTR 154-159	-
Z49770 (Genbank)	Sc	116 NHQ 118	FPEGTN 215-220	LDAIYDVTI 265-273
P26647 (Swissprot)	Ec	72 NHQ 74	FPEGTR 145-150	-
Z49860 (Genbank)	Bn	-	FVEGTR 90-95	VPAIYDMTV 138-146
U89336 (Genbank)	Hs	95 NHQ 97	FPEGTR 168-173	-
U56417 (Genbank)	Hs	103 NHQ 105	FPEGTR 176-181	-
AB005623 (Genbank)	Mm	100 NHQ 102	FPEGTR 173-178	-
Z29518 (Genbank)	Zm	91 NHR 93	FVEGTR 170-175	VPAIYDTTV 218-226

Hs = Homo sapiens, Ce = Caenorabibtis elegans, Ec = Escherichia coli; Sc = Saccharomyces cerevisiae, Bn = Brassica napus, Zm = Zea maize, Mm = Mus musculus

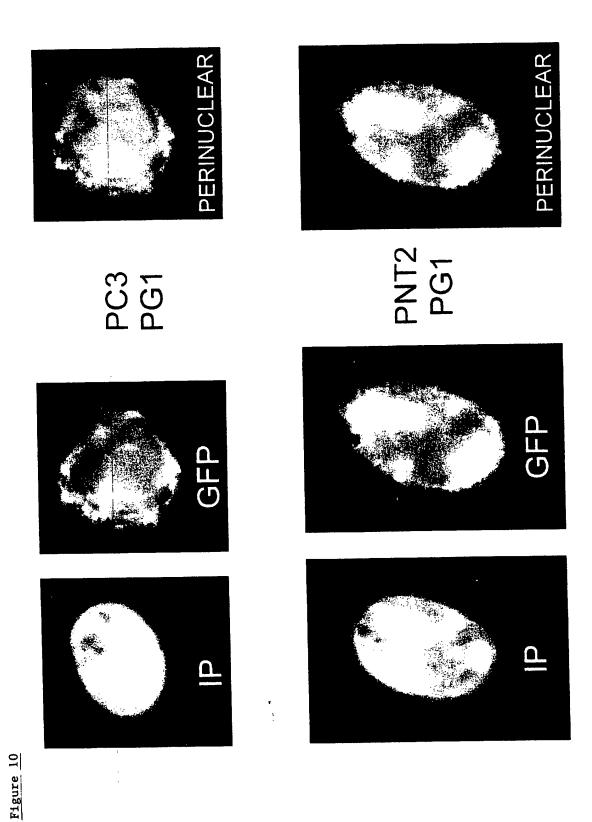
Note: Functional acyl glycerol transferases all contain boxes 1 and 2 and not box 3. Proteins most related to PG1 contain the 3 boxes with a high degree of conservation.

Figure 9

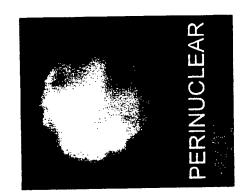
^{- =} pattern absent from protein sequence

11/24

WO 99/32644 PCT/IB98/02133



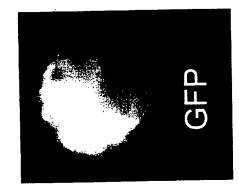
BNSDOCID: <WO___9932644A2_1_>

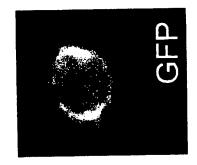


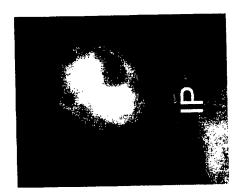
FERINUCLEAR

PC3 PG1 1-4

PNT2 PG1 1-4







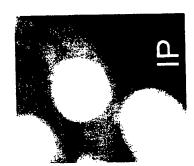
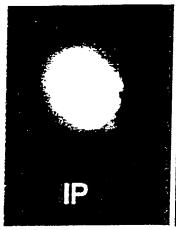
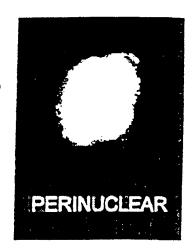


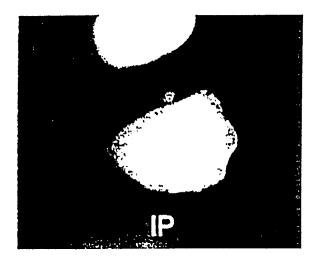
Figure 12

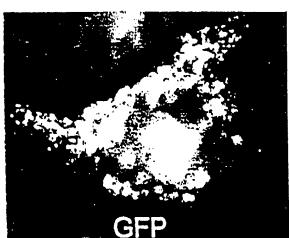




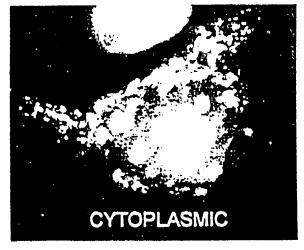
PC3 PG1 1-5

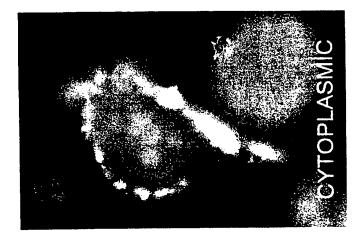


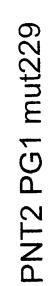




PNT2 PG1 1-5







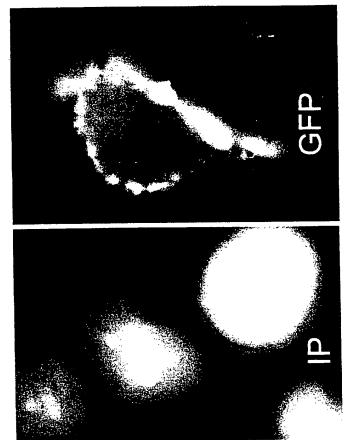


Figure 13

Figure 14

Alternative splicing

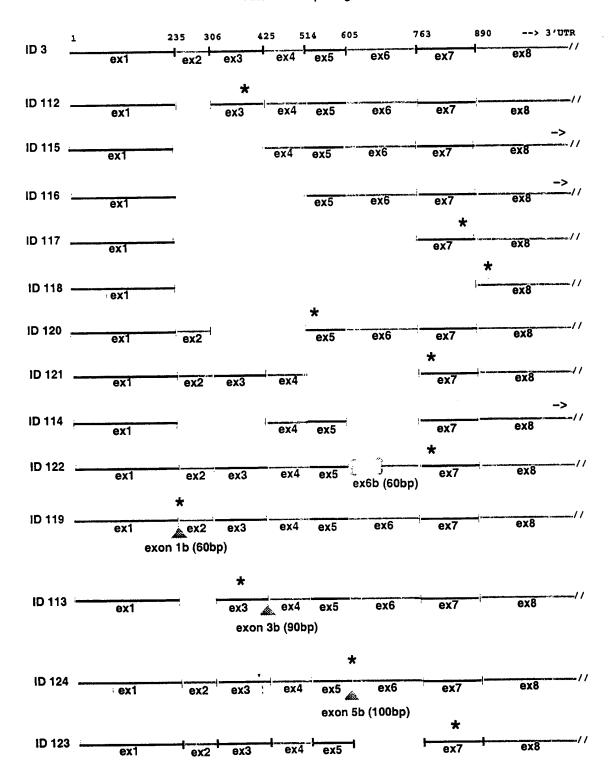


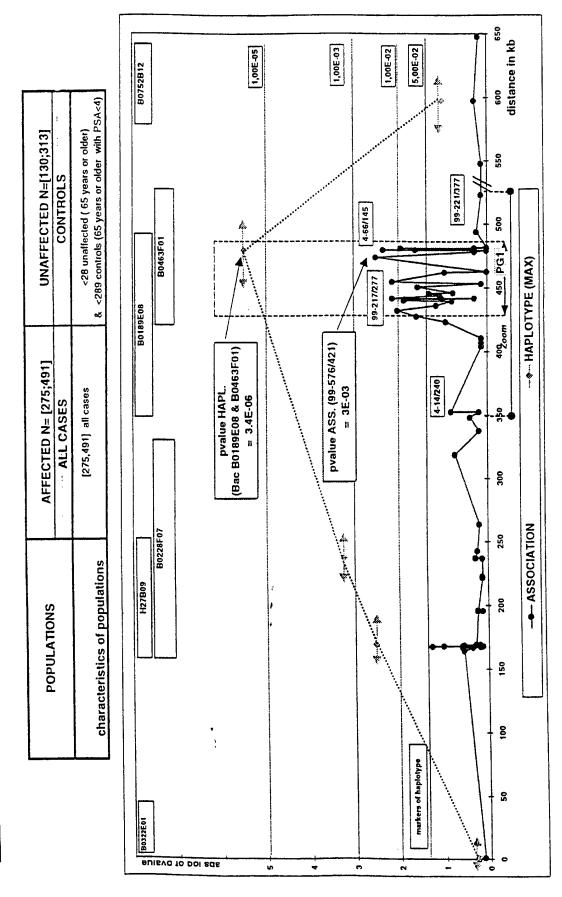
Figure 15 16/24

Combinations of exons of PG1 gene discovered by PCR with primers specific for exon borders

		e.				FCG	JWB																							
Primer	Clones	prostate	PNT1A	PNTIB	PNT2	LnCaPFCG	LNCaPJMB	CoHPV	Du 145	<u>ლ</u>	ECP5	ECP6	ECP7	ECP8	ECP9	ECP10	ECP11	ECP12	ECP13	ECP14	ECP15	ECP16	ECP17	ECP18	ECP19	ECP20	ECP21	ECP22	ECP23	ECP24
		-	-		<u> </u>	-22-1	-			OCCODEDES O			*****	[2]		(A)		(E)												
PG1exon13											*		***				Ť													
PG1exon14 PG1exon15	+		****	3.00 -	****			NT		***	+	00.000 -	***	(Boudin	******	2000			20.63		•	•				-	-	-	-	****
PG1exon16	+	+	+	+	+		_	NT NT	-	-	-	-	+	_	-	-	\exists	-	-		-		_	-	-	╗	+	ᄀ		ᅱ
PG1exon17	+	+	+	+	+	+		NT	+	+	-	·	+	-			+	+	+	+	+	+	+	-	-	+	+	-	+	T
PG1exon18	+	+	+	+	+	·		NT	-	+	_		•	$\overline{}$	-	-	-	-			+	$\overline{\cdot}$	-			可	-1	-	- 1	-
PG1exon24						***	**								**	**					***	**	**		•	· ·			2	
PG1exon25	+	+	-	+	+		-	MT	+	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	- 1	·	+	+
PG1exon26						**				***						23						88	2	2						
PG1exon27	-	-	-	+	+	-	-	ΝT	+	-	-	-	-	-	-		-	+	•	-	+	٠		•	•	•	•	-	+	+
PG1exon28	-	-	+		•	-	·	NT	•	$\overline{\cdot}$	•	•	-	-	·	•	•	•	•	•	•	-	-	•	•	·	·	-	-	·
PG1exon35	-	+	+	+	+	+	+	ΝT	+	+	-	·	Ŀ	•	•		+	+	•	+	+	+	-		+	+	+	+	·	
PG1exon36	-	+									9	•						+	4	÷			XX	22				\pm	•	
PG1exon37	-	-	-			-	•	NT	·	·	-	٠	-	Ŀ	•	•	•	•	·	·	٠	-	-	٠	-	٠	٠			
PG1exon38	•	-		·	·	-	٠	NT	•	+	•	٠		-	-	٠	•	-	Ŀ		+	-	·				+	·	-	-
PG1exon46		+				*	٠					+	*	•		**	•		×.					**		•				
PG1exon47				**				338				•	*										***	**	*	*	**	*		
PG1exon48	1 -	Ī -	Ţ -	•	+		-	NT	٠	•	٠	·	+	-		·	•	-	<u> </u>	<u> </u>	Ŀ	-		-	Ŀ	ينا		-		
PG1exon57	14	+				+	*	167		*			*					٠	*	٠					٠		XX			
PG1exon58	Ţ-	-	Ŀ	·	·	+	+	NT	-	•	•	Ŀ	·	-	·	Ŀ	·	·	·	·	Ŀ	<u> </u>	Ŀ	Ŀ	-	-		+	Ŀ	1
PG1exon68	•	+	+	·	+	+	+	NT	+	+	-	_	Ŀ	Ŀ	+	+	Ŀ	ŀ	+	+	Ŀ	Ŀ	+	Ŀ	-	Ŀ	+	<u> </u>	+	+
PG1exon11b	+	+	+	+	+	+	+	NT	+	+	Ŀ	+	Ŀ	+	+	+	+	+	+	+	Ŀ	-	+	·	·	+	+	+	Ŀ	+
PGlexon1b2	+	+	+	+	+	+	+	NT	+	+	ŀ	+	ŀ	±	+	+	+	±	+	+	Ŀ	Ŀ	+	·	ـنــل	+	+	+	-	+
PG1exon1b3					2				**			*				•	٠								,	***	**			
PG1exon1b4		+	Ŀ	+	+	+	+	+	+	<u>l +</u>	Ŀ	Ŀ	+	Ŀ	+	+	·	-	·	+	+	+	+	·	+	+	Ŀ	Ŀ	+	
PG1exon1b5		+	Ŀ	+	+	+	+	NT	+	+	Ŀ	Ŀ	+	۰	<u> </u> -	<u> </u> -	+	+	ļ-	<u> </u>	+	<u> </u> -	+	-	ŀ	+	+	+	+	-
PG1exon1b6							*		•	٠	٠	+	*			\otimes	1	*	*	*	*		2		*	*	*	*	2	22
PG1exon1b7		+	ŀ	+	+	<u> </u>	+	NT	+	+	1+	+	Ŀ	Ŀ	Ŀ	<u>+</u>	<u> -</u>	ŀ	<u> -</u>	ŀ	<u> -</u>	۱÷	ŀ	ŀ	·	Ŀ	Ŀ	<u> </u>	ŀ	
PG1exon1b8	3	Ŀ	<u> </u>	+	<u> -</u>	+	١.	NT	با	ļ .	<u> </u> -	+	<u> -</u>	ŀ	ļ.,	-	٠.	-	ļ ·	-		-	ļ -	ļ -		· ·	<u></u>			
PG1exon3b4	•		٠		*	*	*	10	<u> </u>	+	1	*	1.2	Ľ	12	Ŀ	*	*	13			Ł		! *	! *	•	*	*	***	
PG1exon3b5	_		Ŀ	+	<u> +</u>	<u> .</u>	<u> </u>	NT	بنا	با	ŀ	٠	<u> </u> -	<u> </u> -	-	ļ -	Ŀ	<u> -</u>	ښل	ļ٠	<u> </u> -	ŀ	-	-	ļ -	-	٠.	-	ļ	
PG1exon3b6	_	*			+	+	•	36 7		*	×		Ŀ	-	1 ±	1 *	*	1.	<u> 1 + </u>	*	1 *	1 *	Į±.	Ŀ	1	<u> </u>	<u> </u>	*		*
PG1exon3b7		+	٠.	+	+	ŀ	+	NT	Ŀ	+	·	+	ļ٠	·	ŀ	ļ٠	ŀ	Ŀ	ŀ	<u> ·</u>	+	+	ŀ	1-	+	·	ŀ	ŀ	ا نــ	۲
PG1exon3b8		<u> </u>	ŀ	·	•	-	ŀ	NT	<u> •</u>	Ŀ	ŀ	Ŀ	<u> -</u>	<u> •</u>	٠	ا نا	┶	ļ-	ļ٠	┵	<u>ا</u> نا	ļ٠	ŀ	<u> </u> -	↓ -	ļ -	ŀ	ļ٠	┶	لنإ
PG1exon5b		+	Ŀ	ŀ	ŀ	ŀ	+	NT	Ŀ	+	نا	+	ļ-	1-	ļ٠	Į∸	+	+	<u> </u> -	ŀ	ŀ	<u> -</u>	 -	ļ-	ا ن	<u> -</u>	<u> </u> -	ŀ	<u> </u> -	ــــــــــــــــــــــــــــــــــــــ
PG1exon5b		+	1 -	+	+	+	+	NT	+	<u>↓-</u>	ا	┵	↓-	ļ٠	1-	ا نـــٰ	ļ٠	ŀ	↓ :	ļ٠	ļ٠	+	ļ٠	ŀ	╀	٠.	۲	+-	┿	ᆣ
PG1exon5b		<u> </u> :	1.	1.	<u> +</u>	<u> </u>	<u> </u> -	NT	- ا	۰	ļ٠	<u> </u> -	<u> </u> :	<u> </u>	<u> -</u>	<u> </u> -	ـــــــــــــــــــــــــــــــــــــــ		ļ -	1 -	1 -	1 -	ļ :	-	1:	-	٠.	1.	4	4
PG1exon56						**	1 *	•	*				-	-			÷	1 *	1 *	1			****	Ŀ	•	•	*		-	
PG1exon46			4				12	-	Ŀ	*	*						٠	Į±	4	¥.	*				•	*			14	**
PG1exon36	-				12						} →	12		-			•	! *	1 *		1+			Ŀ		*			1.5	
PG1exon26	_	4	***		12		12		12	-	1.0	1 *	-		*	+	14	1*	*	*	12	**		-	-	¥±	! *	1*	+ *	#
PG1exon16	b 🐘		88			1 -	+)er	1 +	•		•	*		* *	4 *		*			*	*			•	•	•	•	•	

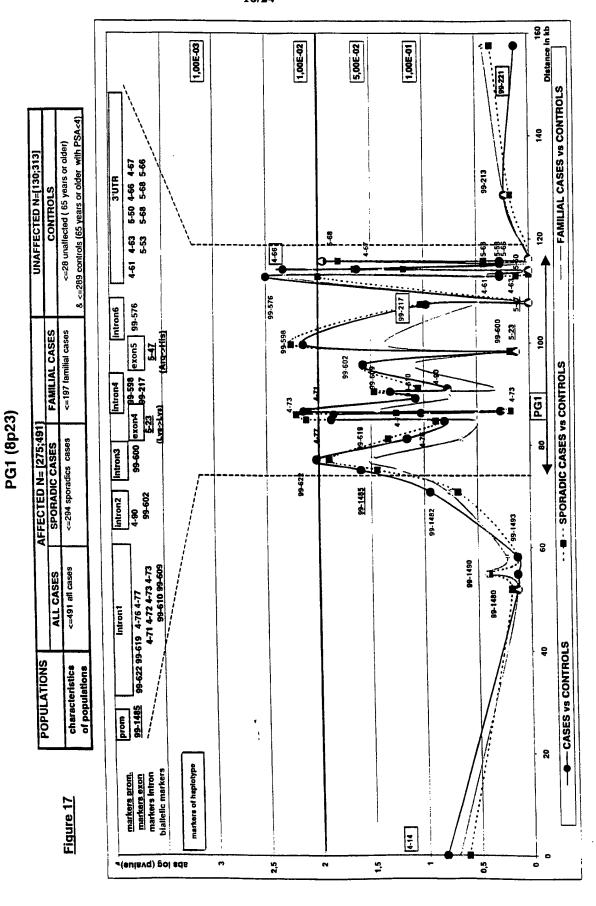
^[+] atternative splicing form with combination of exons 13478 instead of 1345678

HAPLOTYPE FREQUENCY ANALYSIS (chrom 8p23) **ASSOCIATION STUDIES**



18/24

ASSOCIATION STUDIES



ASSOCIATION STUDIES

PG1 (8p23)

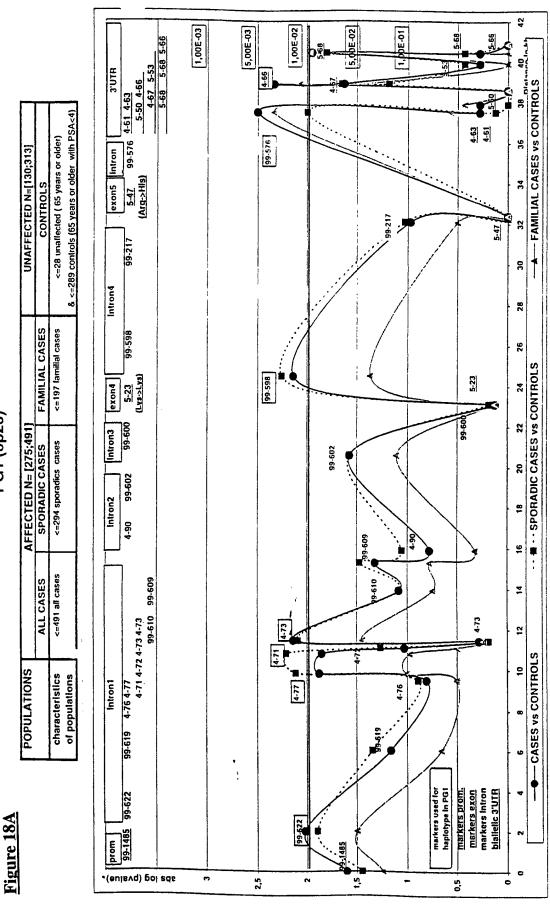


Figure 18B

					abs diff %				
					(fq(cases) -				Attributable
	ž	Bohm	Fred(cas)	Freq(controls)	(fq(controls))	Odd Ratio	Pvalue	Freq(randoms)	Risk
name of markers	5		0.30	0.24	7.4	1,44	2,53E-02	0,29	17,58
99-1485/251	prom	5 6	20,0	0.42	10.1	1.51	9,64E-03	\$ QN	QN
99-622/95	Ē	(3/1	25,0	2,00	8 4	1.37	6.93E-02	QN	QN
99-619/141	in1	5	0,28	0,42	3	1 23	1 575-01	0.41	13,15
4-76/222	ini	G/A	0,43	0,38	c	62,1	0 110	100	10 16
4 77/454	int	2/9	0,34	0,26	7,4	1,43	1,355-02	15,0	01,01
101//	į	Α/G	0.34	0,26	8,3	1,49	1,43E-02	0,28	18,64
4-71/233	Ē .	2	0.36	030	5.7	1,29	9,43E-02	0,31	13,25
4-72/127	Ē	500	85	0.42	9.7	1.48	7,29E-03	0,52	26,76
4-73/134	in1	2/5	70,02	2F12		1 30	8 33F-02	QN	QN
99-610/250	<u>=</u>	G/A	0,43	0,37	2'0	00,1	20 100,0	C.	Ę
300/003 00	Ë	TA	0,37	0,30	7	1,36	4,83E-02	200	2
23/600-66		C/ V	0.20	0.25	4,4	1,25	1,68E-01	0,28	9,32
4-90/283	· INZ	2	210	20.05	7.4	1.44	2,69E-02	QN	QN
99-602/258	in2	A/G	0,50	200		5	7 52F-01	QX	QN
99-600/492	in3	T/A	45.O	\$6,0	20	2.4	7 205-03	Ç	GN
99-598/130	łui	G/A	0,35	0,25	9,2	1,55	20-36-7	200	8 46
776/212	4ri	1/C	0,31	0,28	3,8	1,20	1,0/E-01	0,28	0,,0
33-2111211	92.	9/0	0,27	0,17	9,2	1,72	3,18E-03	0,24	18,40
99-5/6/421	21 17	5 0	0.01	0.00	0,3	1,76	0,527 §	QN	Q
4-61/269	зит	V/5	2,2	0 10	6.2	1.43	4,68E-03	0,24	13,16
4-66/145	3'UTR	5	0,23	60.0	100	1 33	2.39F-02	0.24	10,97
4-67/40	3°UTR	1/C	0,25	0,20	4,3	ł	2,22,2		
				R Toet Fiehe	E Toct Eicher . & ND: Not done .		disease associated allele / not associated allele	of associated allere	

PT4 РТЗ

HAP 10 <160kb> HAP 11 <160kb>

HAP 9 <117kb>

HAP 8 <60kb>

PT2

HAP 6 <160kb> HAP 7 <160kb>

HAP 5 <60kb>

HAPLOTYPE FREQUENCY ANALYSIS

POPULATIONS	AFFECTED	UNAFFECTED
sample sizes	CASES (n=491)	CONTROLS (n=317)
characteristics	294 sporadic cases	28 unaffected (65 years or older)
of populations	+ 197 familial cases	+ 289 controls (65 years or older with PSA<4)

		_								
	4-14/240	99-217/277	4-66/145	99-221/377	haplotype	type				
		lin4	3'UTR		frequencies	ncies				
nks	<100kt	٨	<17kb>	<43kb>						
rols)	481 vs 305	481 vs 302	481 vs 300	481 vs 303						
controls)	65.7/62,1 (C)	31,3/27,5 (T)	25,1/19 (C)	42,7/42,91 (A)					Č	
controls)	3,6	3,8	6,2	0	cases	controls	ppo O	S-I-S	Pvalue	ø.
	1,47E-01	1,07E-01	4,68E-03	7,52E-01			ratio			
cases	5,84E-01	6,55E-01	2,54E-01	5,84E-01						
controls	4,80E-01	2,21E-01	3,71E-01	2,54E-01						
454 300 2007			O	4	0.116	0.067	1.83	9.85	(1.7e-03)	:
451 VS 237		+	· U		0.243	0.183	1.43	7.49	(6.2e-03)	**
451 VS 230	(•	· c		0.182	0.130	1.49	7.18	(7.3e-03)	**
452 VS 299	، د	۲	•		0.217	0.188	1.20	1.88	(1.7e-01)	
4/9 vs 302	د	- +		4	0.155	0.132	1.20	1.54	(2.1e-01)	٠
476 vs 300	,	-		: ບ	0.373	0.346	1.12	1.16	(2.7e-01)	*
476 vs 303	، اد		٠		0.095	0.042	2.39	14.62		****
447 vs 297	د	۲) C	: ∢	0.117	0.065	1.93	11.33	(7.3e-04)	***
446 vs 294	,	- +) C		0.178	0.125	1.53	7.80	(5.2e-03)	**
450 VS 296	، د		•	٨	0.114	0.089	1.32	2.44	(1.1e-01)	
4/4 vs 300	ر	-			900	0 000	2.18	21 59	(3.40-06)	****
445 vs 294	ပ	۲	ပ	A	0.093	0.032	3.5	21:22	22.22	

HAPLOTYPE FREQUENCY ANALYSIS PG1 (8p23)

markers	4-14/240	99-217/277 in4	4-66/145 3'UTR	99-221/377
ot haplotype Max	ပ	⊢	ပ	A
distance between mks	<100kb>		<17kb>	<43kb>

				- 44	S-ido	<u>م</u>	P value
: (sample sizes	hapk freque	haplotype irequencies	odd	2		
F54	cases vs	cases	controls				
	AAE ve 294	0.095	0.032	3,18	21,59	3,40E-06	*****
cases vs controls	171 16 201	0.105	0,032	3,56	20,91	4,60E-06	*****
cases (<=65 years) vs controls	271 vs 294	0.079	0,032	2,60	12,13	4,80E-04	****
cases (>65 years) vs connois	107 CA 1 177	960 0	0.032	3,23	19,73	8,60E-06	****
sporadic cases vs controls	700 NS 234	0,035	0.032	3,20	12,04	5,00E-04	****
sporadic cases (<=65 years) vs controls	425 VS 234	0,085	0,032	2,82	12,75	3,50E-04	****
sporadic cases (>65 years) vs controls		0.062	0.032	2,00	2,70	9,40E-02	**
informative sporadic cases vs controls	I	900.0	0.032	3,32	18,33	1,80E-05	****
familial cases vs controls	1/9 VS 284	0,030	0.032	3,83	17,98	2,20E-05	*****
familial cases (<=65 years) vs controls	86 VS 294	0.075	0.032	2,48	62'9	1,00E-02	34
familial cases (>65 years) vs connos	70.00.07	0.123	0.032	4.26	21,33	3,70E-06	*****
familial cases (>=3 caP) vs controls	19 VS 234	23160					

HAPLOTYPE FREQUENCY ANALYSIS (PG1)

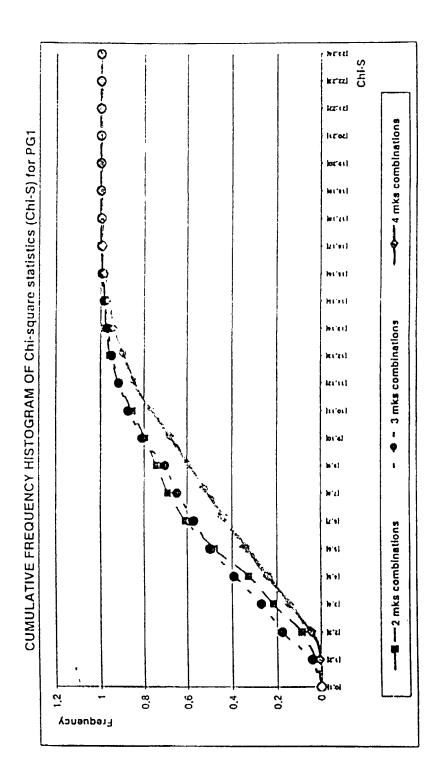
			99-622/95	4-77/151	4-71/233	4-73/134	99-598/130	99-575/421	4-66/145	hapi	haplotype				
Marke	Markers in PG1	<u>a</u>	5	ည	ΑG	g/o	A/G	כעפ	5	·					
				r.	1		in4	gui	3.UTR	frequ	frequencies				
stre (cas	stre (cases vs controls)	(S)	336 vs 108	363 vs 173	336 vs 130	352 vs 129	347 vs 126	355 vs 129	456 vs 306						
allelic frequency % (cases / controls)	y % (cases / c	controls)	52/42 (G)	3476 (G)	34/26 (A)	52/42 (G)	35/25 (G)	(5) (1//2	25/19 (C)		,,				
	alkilc frequency % (randoms)	loms)	QN	31 (G)	28 (A)	52 (G)	ON	24 (G)	24 (C)						
	diff freq. all. % (cases-controls)	ntrois)	10,1	7.4	8,3	9.7	9.2	9.2	Z9	cases	controls	рро	Attributable		:
			9,64E-03	1,35E-02	1.43E-02	7.29E-03	7,29E-03	3,18E-03	4.68E-03			Ratio	RISK %	7 2000)	rvaide
	ovalue (cases vs controls)	ols)	:	:	:	:	:	:	:					ea sacan	(success)
	Odd Ratio		15.1	1,43	1.49	1.48	1,55	1.72	1,43						
	Arributable Risk %		QN	18,16	18.64	26,76	Q	8.46	13,16						
Hardy	\$0500	\$ 0.2	7.52E-01	7,52E-01	5,84E-01	7,52E-01	7,52E-01	7,52E-01	3,43E-01						
Weindeberg Disequilibrium	controls	rois	4,39E-01	4,03E-01	1,21E-01	7,52E-01	6.52E-02	7,52E-01	1,29E-01						
haplotype 1	2 MKS	339 vs 167		U					ပ	0,263	0,152	1,99	18,55	(6.79-05)	<u> </u>
парюнуре 2	3 MKS	330 vs 122					ប	g	ပ	0,259	0,147	2,02	Ş	(3.98-04)	<u> </u>
парюнуре 3	4 MKS	312 vs 122			∢		U	ប	ပ	0,259	0.147	2,02	QN	(4.18-04)	:
haplotype 4	S MKS	311 vs 121			∢	g	U	g	υ	0.26	0,148	2.01	QN	(4.80-04)	:
haplotype 5	6 MKS	309 vs 121		g	∢	ŋ	G	ប	ပ	0.258	0,149	2	Ş	(5.39-04)	:
haplotype 6	7 MKS	290 vs 99	g	S	A	g	g	G	ပ	0.255	0,146	2	Ş	(1.60-03)	<u> </u>

Figure 2

SUBSTITUTE SHEET (RULE 26)

Figure 21

combinations # of 4 mks 3876 Comparison of Pvalue between nb of mks for haplotype combinations # of 3 mks 696 (19 mks of PG1) combinations # of 2 mks # of markers GENE PG1



1

```
<110> Cohen, Daniel
     Blumenfeld , Marta
     Chumakov, Ilya
     Bougueleret, Lydie
<120> Prostate cancer gene
<130> PG1
<150> 08/996,306
<151> 1997-12+22
<150> 60/099,658
<151> 1998/09/09
<160> 578
<170> Patent.pm
<210> 1
<211> 56516
<212> DNA
<213> Homo sapiens
<220>
<221> promoter
<222> 1629..1870
<223> identification method Proscan
<221> misc_feature
<222> 1998..2000
<223> potential ATG
<221> exon
<222> 2001..2216
<223> exon1
<221> misc_feature
<222> 2031..2033
<223> ATG
<221> misc_feature
<222> 11694..14332
<223> Tyr Phos
 <221> primer_bind
 <222> 11930..11947
 <223> upstream amplification primer 4-77 SEQ ID42
 <221> allele
 <222> 12057..12103
 <223> polymorphic fragment 4-77 SEQ ID24
 <221> primer_bind
 <222> 12339..12358
 <223> downstream amplification primer 4-77 SEQ ID51, complement
 <221> primer_bind
 <222> 13547..13564
 <223> upstream amplification primer 4-73 SEQ ID64
 <221> allele
 <222> 13657..13703
 <223> polymorphic fragment 4-73 SEQ ID58
 <221> primer_bind
 <222> 13962..13981
 <223> downstream amplification primer 4-73 SEQ ID67, complement
 <221> exon
 <222> 18196..18265
 <223> exon 2
 <221> exon
 <222> 23717..23832
 <223> exon 3
 <221> exon
  <222> 25571..25660
  <223> exon 4
  <221> primer_bind
  <222> 34216..34234
```

```
<223> upstream amplification primer 99-217 SEQ ID43
<221> allele
<222> 34469..34515
<223> polymorphic fragment 99-217 SEQ ID25
<221> primer_bind
<222> 34625..34645
<223> downstream amplification primer 99-217 SEQ ID52, complement
<221> exon
<222> 34669..34759
<223> exon 5
<221> exon
<222> 40688..40846
<223> exon 6
<221> exon
<222> 48070..48193
<223> exon 7
<221> exon
<222> 50182..54523
<223> exon 8
<221> primer_bind
<222> 51149..51168
<223> upstream amplification primer 4-65 SEQ ID65
<221> allele
<222> 51448..51494
<223> polymorphic fragment 4-65 SEQ ID59
<221> primer_bind
<222> 51482..51499
<223> downstream amplification primer 4-65 SEQ ID68, complement
<221> primer_bind
<222> 51596..51613
<223> upstream amplification primer 4-67 SEQ ID44
<221> allele
<222> 51612..51658
<223> polymorphic fragment 4-67 SEQ ID26
 <221> primer_bind
 <222> 51996..52015
 <223> downstream amplification primer 4-67 SEQ ID53, complement
 <221> polyA_signal
 <222> 54445..54450
 <223> AATAAA
 <400> 1
 gtggatctgt gactgttcgc aggaagagag gagcgggagc aggacagaca ataactgata
                                                                      60
 gtcaggagct gggtttggag ataaagaggg aacaagagaa agttaagttc tgtgttttca
                                                                     120
 tggcaaacat tgcacaaaag tttacaactt cgtgactaac agtaatctgg ggtgattcac
                                                                     180
 aacaaattta cacataaaca catatttact gactttatac acagcaatcc taacgtgaac
                                                                     240
 acagaacctg ctttatcttt tcgcacactg ttctagtgta gagatgtctg gtctcagtta
                                                                     300
 aagaaagcat aaggagcatt agttgtgcac actgtccaca cccgtgactt ttttccacca
                                                                     360
 gtactaaacc tagtgcttct tacagtacag ggcaatgaca gccacagaaa gagagaagct
                                                                     420
 cettttactg tgtaatgett cetgetggee tteaaatact tgttacttga gagateteea
                                                                     480
 ttcacctggc tttgtcccca aaggtcatca tctaccaatg atgttgttat ttgatgttaa
                                                                     540
                                                                     600
 tcatgtataa agaaagtagc taccatcctg gccctgatta gaacttccca ctgaaatacc
                                                                     660
 gtcctgccta aaggtagcac aggtttccat tatggtggtg gtggggaggg ggcgggaata
 tatatatata tatatata tatatatat gtaaagcatt cggcattctt ttaaagtaca
                                                                     720
 actatectty aaaagggtta catattaaac catttttacc acagecaaag gggaggagaa
                                                                     780
 agatecaaaa gteetgtgga tetgetttaa cateaataaa acagttatee accettegta
                                                                     840
                                                                     900
 gcttttagtg aaggctacaa aagtatgctt tttatggatt acacatgtgc acgcaactac
 tttaattact acagaaaaaa acgaggctcc ttattaaaaa aaaatcagaa acaagtccaa
                                                                     960
 cagactetga ggaaatgaag caagagtgaa ttetgaaaag gtetaataaa cagtatggaa
                                                                    1020
 atatecttgt gggattgtte tteagetatg cataaacatg taattateat cattactgtg
                                                                    1080
 atggggaaaa acacggaccc taattctgaa acaccctggt agcgagagac gggcaggagg
                                                                    1140
  ggctgctgcg cactcagagc ggaggctgag gaggcggcgt ccccttgcaa aggactggca
                                                                    1200
```

	ggygacactc	ranctacccc	acaacctaga	ccgagctgcc	tacaacctgg	1260
grgagcagar	ctgcaagaat	tagacctccd	ataacqttaa	cacccacttt	ctcactgctc	1320
geeeaggige	catcccggcg	cccadaaact	tataaacaac	aggtgcgcgt	tccaggcagc	1380
taattgtgtg	cttaaacctg	accaggggca	catccaaccc	dadddadcad	aacaagaggc	1440
tccagcgacc	tcctccggcc	accececec	ttcacccagt	tecateaate	gccaccacct	1500
acceggaeee	gtccgcagcc	agcacccacc	addadcatac	acaataacca	gagccgggtt	1560
cccttcccgc	cagcaggtag	ctatactaca	actataaca	caaaccaacc	aatcaagaga	1620
gcccgcgcca	ccgccgaggt	ccycactyca	accycogyce	ccaatcagaa	gacacattat	1680
cgtgttattg	ccgccgaggu	gyaactatgg	atataacaat	aaataaaaat	catogggtgt	1740
tgccgcggag	cccctgccc	cggcaggggg	acgiggegat	cettaettta	caccaaasa	1800
gagcatccct	gagccatcga	tccgggaggg	cegegggtte	ageageaga	cccatacaac	1860
cggcgcacgc	agccccgcac	tegeetaeee	ggeeeeggge	ggeggegegg	caataattaa	1920
tgggggcgga	ggctgggagc	gggtggcggg	egeggeggee	egggeeeggg	cggcgaccgg	1980
ccgcctgctg	gccgcgactg	aggcccggga	ggcgggcggg	gagegeagge	ggageteget	2040
gccgccgagc	tgagaagatg	ctgctgtccc	tggtgctcca	cacgtactcc	atgegetace	2100
tactacccaa	catcatactc	ctgggcacgg	cgcccaccta	cgtgttggcc	tggggggtet	
ggcggctgct	ctccgccttc	ctgcccgccc	gcttctacca	agcgctggac	gaccggctgt	2160
actgcgtcta	ccagagcatg	gtgctcttct	tcttcgagaa	ttacaccggg	gtccaggtga	2220
accacctece	actcccaaat	ctcggcgtcc	acccgagctc	ccgggggcgc	ggacctctcc	2280
geteccecae	agctggcgag	ggtcacccgg	ccggcccggc	ggacccagca	cggagagcac	2340
ataccacctc	cccaccttcc	tctccgcatg	cttcctgccg	ttctgccgag	ategetetet	2400
aggaagctgt	aactacatca	tcctgaggct	acgagtggga	cccgccgccc	ctttccccgc	2460
ccctcaccta	ggtctgatgc	tgcttagcaa	agtgggtgca	gatgcacgtt	ttaaataata	2520
aggcacgcgt	ttagcagttt	ctggcctttg	gtccaaagag	gtggtcatgt	tggaacagat	2580
cagagacatc	tacactccga	agtgcgcttt	tacagtgacc	tcttgaaaca	gaagtacaat	2640
tegatettat	gttctttccc	ctggacaagt	gaaagctggg	cgaagaaatg	aatacatttg	2700
ttaaccutau	aagcctaact	agatacaatt	cttgccaact	ttaactgggc	ttgaatgtgt	2760
gaataateta	ttgtctgatt	actttcttc	tgttactgtt	tctctgtaga	gattggattc	2820
gggtgatteg	cttgagaaac	aaaccataaa	agtggaaggc	cctctttaac	agtaggtatt	2880
tasatatts	taaaaaaaaa	aaaggtgaat	ttttctttta	tttctcagtt	tgaaagaaca	2940
gaagtgtta	tggttattcc	taatgtccac	ctagtcctct	tttacttttc	ttggtagggt	3000
toggetage	tggggaaatg	ggacggtatc	attttgtctt	tttaactttt	ttttttcca	3060
cagggraggea	ctgtttttac	cctataatca	gtcaggtact	atatttagtt	tgcagttgca	3120
cctacagcag	accettgatg	accccaatta	gaagttgttt	ggggggaagg	aactaggaga	3180
etgetgateg	tccatttaaa	gccccagccg	taagtgtctc	cttggaagga	aaaaaagata	3240
ggccagggcc	tcatggtttc	ctagtgttta	acotttaaaa	taggcctcat	ttaaaaattt	3300
ctgttccagg	ggctaatttt	ttccctttat	atggttaactc	caccaagttt	gtctaaatgt	3360
caataattca	tcatgattaa	-ttttaat	ccacatcata	taacaactaa	cctgggatgg	3420
atgattttta	tcatgattaa	gtttttactt	ctattassas	aataattota	tctataacaa	3480
gatataagct	cagaacacaa	agteatteac	ototaaaaa	gatactcca	ttgaaggatt	3540
gttatgttat	ttttgttcaa	agaggacaca	atatyatyca	gaacacacca	tatttctaaa	3600
ttttggtttg	gcaagttctt	attttttaa	atggetgtaa	additagtag	caccaactea	3660
attgcatacc	ttacctgatg	ttcagagatc	cgatttactt	ttenatten	cagcaagcga	3720
ttttgaaaac	atttaatcta	atcattcccc	ccaccgtcty	Licaaaccaa	aggaagegge	3780
atccagcact	aattttcatg	catttatgaa	aggatgcctg	aggaeeetta	agtataatte	3840
aaaattttgt	ttaatgtgtg	ttccttgatg	aagttcttta	ggagtegtag	aacyaactya	3900
ttgcccactg	atcatcaaat	gcaagttatg	aacatttaat	aaaaatttaa	aaccaagagt	
ttcttattcc	· rocatttta	tttttattqt	atggagggga	caaataatta	Littergett	3960
agtaacagag	cagggtattt	tgaatttatt	agggtctttt	tctgcagtct	gggtttcctg	4020
tgtacacaaa	gctacctttc	aatattttt	attgtttctg	ttaagattaa	. atcaatagag	4080
gaataaatag	ctatcttcaa	acataagacc	caaaggaaaa	. agatttatag	tgatgttctg	4140
tcaccttatt	ttttacctgt	gactttgtac	cattaacttt	. gtcactgaga	tgttttgatt	4200
aaaattttta	achtacttt	cttattttat	. taggacactc	: tttttttctt	gaattgtttt	4260
tatcagetti	catttacaaq	gctagtgatg	attctcttgt	: tctgtataaa	gtattgttga	4320
ctcatttctc	aagggagttt	. taqtaattta	. agaggttata	. agtttttaaa	taaaaggttt	4380
attaatttai	t atatattaaa	gaggcatttt	: aaaataaaat	: tttttttaaa	tgacattttt	4440
acacctttc	a actotaggtt	taaaaaataa	ı gtggttcaca	ı gtagttcttç	, cagaagaata	4500
ttttcttt	a catagaattt	ttaagctgaa	ı gagaagtagt	: agtaggtcca	a tgagatttat	4560
gatetgtge	t taacaaataa	acctacttca	: aacaaattta	ı gttggattt	tcttggattc	4620
tagataaat	a cotttttctt	ccccaqtttc	achactttat	: tttcatatgt	: atctctgaga	4680
tagagaaat	a titcagicad	r tactactaaa	a attgttcctt	: ataactcgt1	tatcctttta	4740
aatcettee	a gaatetetea	a ttggtactga	a aactcaaatg	g ggtactttci	t tcaccattta	4800
tttctttag	a ataagtaata	a agaattttat	aagcttttt	atatttcac	g taatttgaga	4860
cacccaag		-	-			

ctattgaaaa	tccagttaag	tctctctact	gtgttgagag	gcattgattc	aagtacctgt	4920
gttactttcc	tgtgctgcca	aaacagatca	cctcaaacta	agcggcttaa	aataatagaa	4980
cttaagttct	cataattcta	gaggccagca	ctttgaaatc	aaggtgtagg	ctcaatttta	5040
ctccctctaa	aggccctagg	gggaatctgt	tcttgtgggt	ttcaacttct	ggtgactggt	5100
ggcattcctt	aacttaaggc	cccatcactt	caacctctgc	cttacagtcc	ttgctgccac	5160
ctcttctatc	tcacatctca	ctctcccttt	ctcttagaag	gatgcttgtc	attgggttta	5220
gaggggaggt	ggatattccg	ggatgatctc	ttcatctcaa	gatccttaat	tataactgca	5280
aagagccttt	ttccaaataa	gaaaacattc	acaggttcca	gggcttagga	tgtggacaca	5340
ttttttgagg	ggctgccctt	cattccccca	caacaatgaa	ctccatagtt	ctgcctattc	5400
agtattttgt	agttatttcg	tagtttaact	tgccttattt	ctttaggtat	ttacgtatta	5460
aagcattttg	gtctctgctt	tctttaacag	agaacctggt	tttctgtaat	aagtttactt	5520
actttcccat	aatcttttag	tttcttattt	acagatttac	cttcacatat	cccttaagta	5580
gaacatttga	ttaactgttt	tattttcgga	acaaatctgc	attctgtata	ataaccaact	5640
tattcatatt	tcggtattct	tttaattctt	atctgattct	gaaattacca	tcttgtgatt	5700
aratatatat	atatatggaa	ataactgaaa	tcttgataaa	ttaaaggtga	tataacttct	5760
aagacaatta	attatgtatg	atgtggtgaa	tatactggtg	tttggtttgt	ttgccactta	5820
aaaaccctat	ctataggata	ggaagtaact	tgaatgtgga	atgcttagag	actcagagta	5880
agagggggta	tatatatcct	tgagctggag	tttaaggaaa	acttatggga	aattaaaagg	5940
agaggeegea	tactgacaga	ggattgcgta	ggactcatga	aaaaqqaatg	aagttacctt	6000
aaagttggug	atcgtgagtt	aacgtgaaac	tagatttatg	ttagtttata	gcctagaatt	6060
ctatectage	aatctagata	tatcctaaat	gttgagatag	ctgcataaac	aataactgta	6120
atcortatoa	taaataatga	caaatctttt	tagcatgttt	tgtgaagctg	ataaatgtta	6180
ataggatata	ttcaaatgtc	agaattettt	tttctttact	tcttttttaa	aaaatttctt	6240
troccocatt	cctatgcaat	acactgaaaa	ctgatcattg	aaatttgtag	gccaaaaaat	6300
taatcaacac	gtaatagatt	aggatttaga	ttttttqaq	tcagggtctt	cttctgtcac	6360
ccaggeteta	gtgcggtggc	accatcatoo	ctcattgcag	ccttgaatgc	ctgggttcaa	6420
atastacta	ggagtagctg	ccgtgccatt	atttctagct	aatttttaaa	agtttttgta	6480
gegatectet	ctttctgtgt	tacccagact	ggtcttgaat	tcctggcctc	aggtgatcct	6540
tetacettaa	cctcccaaag	toctoggatt	acaggtgtga	gccaccatgc	ctagececta	6600
ataaatatto	taattaccga	tttatcttgc	ttaaatcagt	tggtaacact	tggaatttac	6660
ttcacaatat	attttacatt	agtggctctg	actoctaatt	ccccttctc	caaatgctaa	6720
totaguatus	caataaaatg	cacagttctt	aagtttatat	aaaataaaca	ggttttcagt	6780
taacctactt	taagtgtaaa	atagtgtgaa	aaacacaaqa	aagaagataa	agaatttaag	6840
attttgacat	ttctctaata	tocccttaac	ttctccaagg	attcatactt	ttttttgtaa	6900
gacagaatct	cacactgttg	cccaaaccag	aggtgcagtg	gtgcagtctc	cactcactgc	6960
aacctctgcc	cccgggctca	agcggtcctc	ccacctcagc	ctcctgagta	gctgggacta	7020
caddtacada	gcaccatgcc	cagctaattt	tttttttaa	tattttttag	tgggggtaga	7080
gaggagattt	tgccatattg	cccagtctgg	ttttgagctc	ctgggctcaa	gtgatccgtc	7140
cttgatccac	catgcttagc	tgattcatac	tcttaactga	aacattgttc	caagtttctc	7200
agaaacagto	aaggcttttt	atctagagaa	catttataac	tggatctttc	tttgtgtagc	7260
actgattcat	caaactaatc	ctaaactcct	aatgagttaa	atttatattc	tgaatcttgc	7320
totaaaaoca	gccattcatt	agaatgaaac	atgtttactt	agaattggag	aagggagctt	7380
ataagtcatc	tagtctactc	ccttttatga	cacttctaca	ttctttctgc	acttctgcca	7440
aaatgttgco	cagcgtcgtc	tctgatacct	atagtcctaa	caagaatatg	aatcatacct	7500
totatectta	attttactct	tctctactta	tttgccattc	atgtgaagac	cttaaataga	7560
tettaaatte	cttccttcac	tttagctgag	agtgacagga	ctgtgtaggt	gtgggtgtgt	7620
ttctgcattt	gcttatttaa	gcaggataat	aaaaactttt	actataggaa	attaaacatt	7680
teccaatea	atacaattcc	agtctaacac	aattaaatto	tggttaggga	actgcttaac	7740
tractagact	tataggaaaa	tactaaaaaa	atgtaactag	aactctattt	ttacacttta	7800
taaatataa	a cctctgtgaa	caaaccagtt	atttcaggtt	gcatttgtgt	: atagttttt	7860
aatgcctgat	ttttctattt	taaaatcaca	gatgcaatta	tacattcaaa	cactgccaca	7920
atactttga	g aaagttaaag	tttcccctac	tectacacte	gcgtacacctt	: tcctaggtac	7980
atcccagtti	r ggtgtgtaac	tttagattt	ttccaagagc	ttttgagtaa	gtgtttgaat	8040
tataggaagg	ttctttagtt	: aaatgaactt	: cttacagato	: agttttttag	g tacagtagca	8100
cgaaatata	c ctgcatacct	: atggggatag	: ctctgtgcca	i ttacgatgga	a aggcacggga	8160
aaacagcac	t cogtatatac	ctagtttact	ttccctctt	: tgtatattt	g tctgattttg	8220
tagaactaa	t getteteaac	togaatcag	a agttaactt	tcctttact	a ttttctcatt	8280
ttattatgg	t ttottaacta	a gaggttgate	ttagtggttg	g gaccattca:	a tagtaagtaa	8340
tgacttttc	a gtaagggato	tctagaacc	c agatccctta	a attcctgca	a tattcccgtg	8400
totacatto	t tecaggiget	t qtcctgggta	a ccaagggata	a caatgtttg:	a tagacaatgt	8460
acctoccat	t atggaggte	a cattctagt	g tgggaagaca	a aacaataac	a agaaaatgaa	8520
		.	· · · · · · · · ·			

						0500
aatttactgt	gccatgccag (gttgtttagc	ctggtgggtg	agaggtaggg	gtttggaaaa	8580
tcttactgag	caagtgacat	ttgtgtggag	ctctgtaaaa	gggccagctt	ggaaggtaat	8640
gtagtcatcc	aggtgagaaa	tgatggttag	gggagtggaa	agagtggatg	ttaagattga	8700
aaagaattcc	aaatctattt	tagtggtagc	tgatagggct	ttgtgattga	atgtggagga	8760
aaaagaagag	ggtgggttag	taacacactc	agtcgcagtt	agtgagtgct	gctgtgtgca	8820
agtattgttc	tattatgtaa	ataattccat	ctttacaaag	taggcaccat	tcttcctctt	8880
ttacagacaa	ggaaaaggga	acacccatgg	ttcacatctg	tagtagccta	gccaggagtt	8940
tcaccacactt	attttctgaa	gatgctctgc	ctggcaatgt	ggttatattg	gttgaaatga	9000
gagggatta	tttcaaggta	ttcatctagg	aaagacatga	actoccaatt	acaatatagg	9060
gaccccctac	aattagagac	atatttatta	actttgccat	acagaggtaa	agtaactctt	9120
acacaccya	ctttgcttgg	attaataaa	aaggetataa	aaattactto	gagtttttac	9180
taaagtaact	cgtaattaac	atagaatatt	taggeededa	aggitticaa	ttgataacat	9240
tttgaacatg	aggagtttga	acggaacgct	tcaagguuuug	ctasattcto	ccccaattaa	9300
aataaacatg	cgttggtttc	attatagtat	agetttteet	totagaccac	ccctccccac	9360
CETECCALE	egttggtttc	acticagict	caaccccttc	cttttctcca	tctqtcatqa	9420
attagaccgc	tcctctctgg	aactccaact	ttatataata	traacatata	taatactgat	9480
tgttacccca	tctcattgtc	agggtaactt	chatgeata	chacatata	acctettee	9540
ataacattag	catattttaa	tgtatggatc	accectety	caacactgta	accccccgga	9600
gatggcaata	atgggaagaa	tgacttgatt	ttactttttc	ttttaacaaa	aacggcggag	9660
tagtctgggc	acggtgtggc	tcatgcctgt	aatcccagca	ttttgggagg	ccaaggaggg	
tggatcactt	gaggtcaggc	attcgagacc	agtctggcca	acattgtgaa	accccatctc	9720
taccaaaaaa	atacaaacac	ttactgggca	tggtggtgtg	tgcctgtagt	cctagctact	9780
caggaggctg	aggtgggaga	atcacttgaa	catgggaggt	agaggctcca	gcttgggcga	9840
cagagtgaga	ccctgtctca	aaagaaaaaa	aaggtaaaag	ggccaggtgc	ggaggctcac	9900
gctggtaatc	caagcacttt	gggaggctga	ggcaatggat	cacctgaggt	cgggagttcg	9960
agatcagect	gaccaacatg	gagaaacccc	ttctctacta	aaaatacaaa	attagccggg	10020
cataataata	cctacctata	atctaagcta	catgggaggc	tgaggcagga	gaatcacttg	10080
aacccaggag	acagaggttg	tggtgagcca	agatggcacc	attgcactcc	cgactgggca	10140
acaagagcga	aattccgtct	caaaacaaac	aaacaaacaa	aacaaaacag	agagaaaagg	10200
cagagtactc	tagggaattc	tagtctgtgt	ttctgtggaa	atgtatatga	atctcacttt	10260
taaqqqatqq	agatttttga	atggcataac	tagttgataa	gttttgctct	aacagggtac	10320
ccaagggatgg	tgagtccgat	tcattctttc	cttaaataga	tgaaggagga	agaaacatga	10380
ctacaccctc	aagagtaagg	cagaatgagc	aaagtcagag	aagttaaaaa	agaattctca	10440
cccaccccc	agtgcagaga	aaccttggtt	tagttgtgaa	tcaaaaccag	tactttttgt	10500
asttttaa	cctatgcaat	tetecaaggt	tttatgttgt	ttcttctatt	tctctgtagg	10560
aacccccgag	caaaacccca	aataanaaan	tottacttoa	agattttaga	gtacttattt	10620
caccayaaac	gtaagtgata	tttagaagac	gactttactg	coctcctcca	acttagcata	10680
gigialaagi	gggcggaaag	asaggaggag	gatggtacct	ggaaaggaga	gtcatgttaa	10740
agaattecag	catattaagt	addggdgggc	tactottasa	andtoraato	ttctagactg	10800
gtcccagcca	catattaagt	gctaaccacc	caccyccada	aggegtaceg	aaagatgtca	10860
acaaaataca	tagtctctac	totaaagtaa	tagttgagaa	attaaagaaa	traacaaara	10920
cttaaaagaa	aacttgaata	tatyctyaya	atannatan	attadagada	agatagaaga	10980
actgaggaaa	taaaggagga	atacaactgt	giccaaaiga	atacccaact	taacccacc	11040
tgttgcatat	gtaagcaggt	ggttcaccta	adayıtyyat	gtaacgtagt	otoggedage	11100
tcttggtgca	cttacatatt	gcattgcttc	cgggcttaat	ttgtgttcat	atayyaataa	11160
attttttgtt	ggtttttaat	tttactcctt	gtaatteegt	ggttgatatt	Cadagugada	11220
aaaattacat	aagcttctaa	tatatgagaa	gtcttctcac	ttgacatttt	ctatttggaa	11220
tttttgcaga	a gagtagtttt	gtcacagtca	aaagattttg	ggatettgea	gtgagaaacc	11340
taggtgtaat	tcctatttct	ctgccattcc	gtatgtcatc	tggattaagt	gtcaacttct	
cagtctcaag	g attctcgtcc	ttaaatggaa	tactttttgt	catgctattt	tgaagacaaa	11400
atgagataai	acgtgaaact	gcctagctca	. gtgaatggta	catcatagat	actcagaaaa	11460
aacacaccct	t ctaaaataag	aacagtacca	. aaagacagga	. tgtaaaataa	gggcagtacc	11520
aaaagacaca	a tocatoctoa	gtgtatgaga	. aagaactttg	tggccttctt	gggtggcaca	11580
ggccatggc	a gttccacagc	atgacgtggt	tgctgtgggt	ggtagagcag	acatgccgct	11640
ccccqtcac	t acctaacttt	gatgettget	: ttcttcagct	gagaggacgc	: agctgtgata	11700
tgaaggtet	r gtgtgtacag	tcgtgacctc	: acatttccaa	ı tttcctgctg	gcagaaccca	11760
cartetaca	a cotacoaoca	ccagagttga	cataaaacaa	r acagcataca	a gaggettgta	11820
acatectte	t ggaaaacact	gtgtaagctt	tcaqtqcqaa	ı taaacatgat	: cagtggcaag	11880
ttctattaa	a totagtetge	: aagcatcctg	, attttactgg	, gcaagactat	gttgatttac	11940
addoddota	a tgattccatg	gatagecea	tactactatt	ttcacaaatt	tcacaagaca	12000
ttcttactq	a aagattacco	tottcttato	g atactgctgc	c ccttttagct	t tcatttgctg	12060
ttcacacta	a acttggagac	tacagtcagt	cagagaactt	getaggeca	ctctcaggtt	12120
attetttea	t tectgateat	cctcaaaatt	ttgaaaaag	a aattgtaaaa	a attacatcag	12180
20000000			_	-		

tlagaacter gtggagtga attycactat geagggat aatteaacat agaagtace titgaatgat gtcaagaga attetgatti ataacaggat atgaatacat agaatgat (gtaaagat gtcaagaga attetgatti ataacaggat atgaatgat accttecat 12360 ggtcgaagat gtcaaagaga attetgatti ataacaggat atgaatgat accttecat 12360 ggtcgaagat gtaaaaatt tgutttaag gcaacagat tittatcag cttigaaat 12540 aacattigt ggaaagacag aagattataca ggattaata tggtgaata tatggaaaga 12540 ggtggagaa cagaagacaga taagacaga ggaggagaa gaacattigt ggaagacaga taagacaga ggaggagaa tagaacaga ggaattecta chcaatactga ggaggaaat aagacagaa ggaattecta chcaatactga ggaggaaat aagaagaagaa ggaagacaga ataagacaga ggaatcatta abcagacaa tacagaga accectgagaa gaacacaga cacacacata tacagacaaa tattgagaaaga cagaattga attaggaaaga cacacacat tetgtigga attagtaaga tatatggaaaga cacacacatactga ggaagacaa taacagaga gaacacaga cacacacaaa tactgaa acceggaacaa gcaacacaa cacacaaa tactgaa acceggaaacaa gcaacacaa cacacaaa tacacaaa tacacacaa gaacacaca cacacacaa attagaaga gaacacaca cacacacaa acceggaaa cacacacaca tacacacaa gaacacaca cacacacaa acceggaa accedgaga accedgaga accedgaga accedgaga gaacacaca tacacacaga ggaacaca acceggaa gaacacaca tacacacag ggaacacaga gaacacacag ggaacacacag ggaacacaa acceggaaacacaa cacagagaa gacacaga gaacacacaga ggaacacaca tacacaga ggaacacaga gaacacacaga gaacacacaga ggaacacacaga gaacacacaga gaacacacaga ggaacacacaga ggaacacacaa tacacaga ggaacacaaga accacagaga accacagagagacacaacaa accacagagagacacaacaacaa gcacagacacaacaacaa gcacagacacaacaacaa gcacacaca							
ttagaaqatt gicaaqagaga atticqaatta gaaaqgaga atgataqaga aactitecat tiggaaqata gicaaqagaga ttittaata gigttagata actitecat gigttoqaaqat tigtaaaaatt tigttitaagi gcaaacagti tittattaag cittigaaa aactitigaga aagattataa gigttaatta taggaatta tataqaacat tittaataa gigttaatta taggaaqat tidaaaattict tooqaattaga agaattatta tidagaatta tatigaaatta taggaaqaa taaaqaaqa tittagaaaat cacatattot tittigaa aaaaggaaa taaagaaqaa titaggaagaa taaagaaqaa cacatattiga giggagaaat taaaqaaqaa gigaaggaaatta cacatattaga gigaaggaaat taaagaaqaa taggaagagaa taaagaaqaa cacacacat tooqaaaataa agaaaaagaa aacacacacat tooqaaaaataa agaaaagaa aacacacacacacacacacacacacac	caacatataq	gontatgtee	ttgagaagca	gcagttaatt	acctaaacac	agcaagtacc	12240
ttggaatgat gtcaaagaga attctgattt ataacagget tttattatcag cetttgaaaat 12420 ggettgaatgat tydaaaaatt tgttttaatg gcaaacagtt tttattag cetttgaaaat 12420 gacttgaata tstgaagaagat gadttatat ggettgaatta tatggeaatt 12540 22540	ttagaactct	gtggagttga a	attgcactat	gcaagggatc	aagtaacaat	aaaattatga	
gstcgaagat tgtaaaaatt tgttttaags gaaacagtt tttattaat tggtastta 12400 gacttgcata aatctggaga aagattatca ggattattaat tgtgtaatta 12500 gstgggagaa cgaagcacatt tttagaacat cacatattct tctgtttgga cagaccact 12500 gstgggagaa cgaagttca gttagttcag tatggcaaag aaaaggcaaa taaagacaga 12600 gstgggagaa cgaagttca gttagttcag tatggcaaag aaaaggcaaa taaagacaga 12600 gstgggagaat cacatattga tgaaacatta agaacataa agcttcaacttgcag gaactcgat tgaaacaaa ttctggaatt acaatttg ggaagtagta 12720 gcaagaattga cacatattga tgaaacaat tctggattc acacattgtt cacaattgttgga acataccaca ttcgttgga attagtgaag ttcagagata tcaatacaca 12780 gcaagaatcaga cacctgagaa tgggagagat tcaatagaag gaacctgatt 12840 gcaaatcaga cacctgaga tgggtagct cagactatac cacaacaca agctaagaaa 13020 aaactgttgga accaaattca ttatttacat tttcaacaag cacaacaca aactacacaa 12990 aaaactgttgga accaaattca ttatttacat tttcaacaag aggtagtggggggaag tggggggaag ggaagagga ggaacctga cacaaataga gaaactgga tgttcatat tctcttaaaa aatctgctcc taatggggg tattcaactg 131020 gaggggatatg tctcaatagagaf gttgtgtac ggtactgaga aggsttgggg tgggggggggggggggggggggg	tragaatgat	gtcaagagga	attctgattt	ataacaggct	atgaatgagt	acctttccat	
gacttygata aatctggaga aagattatca ggatttaata tggtgaatta tatggcatgt 12540 ccaactagaa agaattttt tgcacattat ttetetyttgga cagacacatt 12540 ccaactagaa agaattttt tgcacattat ttetacattag gttcaaaata cctaatgcat 12600 gttgggagaa ctgaagttca gttagttag tatgagaaag aaagcaaa taaagcaga 12760 ctacttgcag gatcotcaag taagccattg acgtggaaat taatagtttg ggaagtagta 12720 ggcagaaca ggaattccta ctcatcaaaa ttctggatta atacaagag gaacctgatt 12780 accoggaaca ggaattccta ctcatcaaaa ttctggatta atacaagag gaacctgatt 12840 atagaccatct tctgttggta atttggaga ttatgtggt cacacttct caaatattt 12900 gcaaatcaga catcaccat atcagcacaa gctaatagca cacactctct caaatattt 12900 caaattgga accacatta atcagcacaa gctaatagca cacctcattga atcaacata 12900 caactgtgga accaaattca ttattacaa ttctacacaa acccaaacaa agctaagaaa 13020 acctgagat tgttctact tctcttaaaa aatctgctcc taa:ggtgga tcatattaat 13080 gaaagttga gtctgtgat ggttggtat gtgtgtgat tgtacacaga gacacctttg 13140 acactgatgga cctcqagttg aggtagcctga cgaaataga aggttggat gctcqaga tctcqagtta gagaacaga cacgtggagaacca cccatgatag gtctgtgat gtgtgtgat tgtacacaga gacacctttg 13200 gggggaacca cacaagatgg gttgtgat gtgtgtgat tgacagtgga tgagacattc 13200 gggggaacca cacatgatg ggttgtcatg cttttaaaa acaccaga gaacacctttg 13200 accgaaacaa gccttcaga tacattctga gaaaaacgaca cacacgtga gagacatct 13200 accgaacaa ttgatcaaga tacattctga gaaaacgacaa agaacaccttg 13200 acgaaccaaga agacttccaa tacatcgga gaaaacacaga gaacaccttg 13200 acgaaccaaa ttgatcaaga tacatctgg agaaaaccaga gagacgatac 13200 acgaaccaaa ttgatcaaa tacaatca ggagaactt tcttaagag gaacactt ctgaacatt 13200 acgaacaaa ttgatcaaa acacactag agaacacga ttcatcatga tgaacatt ctgaacaca accatgagaa accacttga gagacatt tcttaaaaaa aaaactgga atcaacttga gagacatt tcttaaaaaa acaacataga tgataagaca ttgaacata agaacatga ttcaacaaga acaccttga tacacaaga acaccttga acacac	ggtcgaagat	totaaaaatt	tgttttaagt	gcaaacagtt	ttttattcag	ctttgaaaat	12420
aaaaatttgg ggaaagcaag tittagaacat cacatattot titgittgga caagaccatt 12500 ggtgggagaa agaaatttitt titgacaattat titacaattaf gittaaaatt cicaatgcat 12600 ggtgggagaat capaagttca gitagttcag tatggcaaag aaaaggcaaa taaagacaga 12600 ggtgggagaat caatatcga tagaaagata agaacataa agacttcac cacattcacca tagaaagacaga 2720 ggcaggaatt caatatcga tagaaagata agaacataa agacttcaca cacaattaca 12780 acaggaacacatc totgitggta attiggtaga tiatgtggt cacactcti cacaatacaca 12800 gcaaaacaga accacacata atcaggacaa gotaatagaa tcatcotga accagaagaa 12600 gcaaatcaga accacacata atcaggaca gotaatagaa tcatcotga accacatata 12900 tacaggaca cocciggaga tgggtagci cagattac caccaaaca agotaagaaa aactgttgga accaaattca titattacat titcaacaagag gaaccagaga tacacacata 13000 gaaaagttga tgittcatci titcitaaaa aatcgctc taagggggatag tacaagagaa gidaatagaa 2820 gggggatag atcaagagaa gidaaccaga cagaacagag acaccttgi accacagaga accagaga accacttga gggggataga gagaagaga caagttgata tigacaagaga gagaagaacaa 2620 accaggga chotcagat gagaagaga caagttgat tigacaagag tagaagaga caagttgat tigacaagag tigacagaga acaagaga gagaagaacaa 13000 gggggaacaa tigactcaag ggtgtgccc cagaaaagaga gagaagaga gagaagaga acaagaga gagaagaacaa agagagaacaa agaagaacaa agactcaaga gggggaaacta tigacaagaga gagaagaga gagaagaga gagaagaga gagaaga	gacttgcata	aatctggaga	aagattatca	ggatttaata	tggtgaatta	tatggcatgt	
ccaactagaa agaattttt tgacattat tttacattag gitcaaaat cotaagcaat 12660 ctacttgoag gatcotcaag baagccattg acgiggaagta taatagcaag 212760 graggagaat caattactga tgaaagata agaacataa agcotcoat cacaattcc accoggaaca ggaattcca cicatcaaa ttcggatta atacaggag gaacctgatt 12780 gcaaatcaga catcaccatt atcagcacaa gctaataga cacattctic caaatatti 12900 gcaaatcaga catcaccatt atcagcacaa gctaatagaa cacattctic caaatatti 12900 ctacaaggaca cocctgagaa tgggtagac caagtttaa cacacaa agctaagaaa 12900 ttacaaggaca cocctgagaa tgggtagac caagtttaa gaaacgttga tgttctatct tctcttaaaa aatctgctcc taacggaga tacatcaat 13080 gaagggatag atcagaagat gtctgtgatc gtgtcctga cagcttuaacaa gagtragga tacaattaat 13080 acatgttgga accaaattca ttattacat tttcacaag accacag acaaccttg 13200 gaaggggatag atcagaagat gtctgtgatc gtgtcctga gagcacaga acacctttg 24282 gagggaacta dctccaatt agaggaagac caagttgat tgatcaggg gagaacttc 13200 accacagaga cccatgattg ggttgtcctg cttttaaaa cacacagg acaaccttt 13200 accacaagga ccctcagattg ggttgtcctg cttttaaaa cacactgga gagacattc 13200 accacaagga ccctcagattg ggttgtcctg ctttttaaag caaccttgg gagacattc 13200 acagaacaca cccatgattg ggttgtcctg ctttttaaag caacctgga gagacattc 13200 acagaacaca ttgttctaat tctacctatg ctgatacga gggttaga gagagtgat 13200 acagaacaa agctttgaa ttatacaaa aaacattgga agcacatt gtggacaat 13300 atggctgtcaa attgatcaaa ttatacaaa aaacattgga agcacatt gtggacaaa agcttttaa aggaacat tgttgaacaa caacttgga agcacatt gtggacaaa aggacctccc attgttgtaa tgttgacaa acaacttga aggacaatt 13200 aggaaaaatag aggaccac acacactga acaagttt tcgtagaacag aggacacta ttgttgtaa tgttgacaa acaactga acaactgaa aggacacaa aggacacaa aggacacaa aggagacaa agagacaa agagacaa agagacaaa agagacaaa agagacaaa agagacaaa acaactgaa acaactgaa acaactgaa acaactgaa acaactgaa acaactgaa acaactgaa acaactgaa agagacaa agagaacaa ttaaaaaa agagacatt ttaaacaaa agagagaca agagacaa acaactgaa agagacaaca	aaacatttgt	ggaaagcaag	tttagaacat	cacatattct	tctgtttgga	cagaccactt	
getggagaac cugaagttca gitagitcag tatggcaaag aaaggcaaa taaagacaga 12700 gecaggaatt caatatcag tagaaagatt agaaacatta agaacataa agcettcac cacaattcc 12780 gecaggaatt caatatcga tgaaagatt agaaacataa agcettcac cacaattcc 12780 gecaggaatt caatatcga tgaaagatt agaaacataa agcettcac cacaattcac 12840 geaatgacac ggaattcca cacaaattag tatggtgt cacactct tetytiggte attiggtaga tatatgggt cacacattcat 12900 tacacaggac accacatta atcagcacaa getaatagaa teattcetgga atcatcacat 12900 tacacaggaca cecetggaga tgggtagect cacacttac caccaaaca agctaagaaa 13020 gaaagttgg gigttatch tetettaaaa atctgctcc tacagggggat gigttactat tetetaaaa aatcgctcc tacagggggat gigttactac tetetaaaa tacgccc agaaaatgga agctigggg tetaatcaca 13020 gaaagttgg tgetcatact tetetaaaa atcgccc tacagggggattg tacagaagat getggaac cacatagaag aggtigggg tacacacaga gaacctgga caccacagag agaactcaga gagaagagac cacattgg ggggaattag detagaagaga cagtiggag tgaacaggag accatagaagag gggggaaaca gcaatagaagaaga ggtigggaaca gagaacacaga gagaacacaga agaacacaga gagaggaaca accatagaagag ggaggaaca gagaagaaga gagaagaaca 13200 accacaacag gggggaaca gactacaaga ggtigggaaca gagaacacaga gagaacacaga gagaacacaga accatagaaaga ggaagaaca agactacaga ggaggaaca cacatattgg cacatagaaga gagattcaa gaaaagaca tacatacaga gagaacacaa atcacacaa gagaacaca accacacacaga tacaacaca ggaacacaca agaacacaa accacacaca	ccaactagaa	agaattttt	tgcacattat	tttacattag	gttcaaaatt	cctaatgcat	
gcaggaatt caatattota tyaaaaagatt agaacatta agacttcaat cacaattoc 12780 accoggaaca ggaattocta ctcatcaaaa ttctgcattc atacaagag ggaactcgatt 1280 accoggaaca ggaattocta ctcatcaaaa ttctgcattc atacaagag gaacctgatt 1280 accoggaaca cacaccatt accaatcact accaatcact 12900 gcaaatcaga catcaccatt atcagtagaa ttatatgggt cacactcacta 12900 ttacaggaca accaatca tacagacaa gcaatagaa cacaccacaa agctaagaa 13020 acctgtgga accaaatca ttattacat tttacacaaga caccacaaca agctaagaa 13020 acacgttgga tyttctatct tctcttaaaa aatctgctc taatggggg atcaatcaca 13080 gaaacgttga tyttctatct tctcttaaaa aatctgctc taatggggg atcaatcaca 13080 acacgacactg tctcagtg gaaccacg gaacaccag gggggatag atcaagaaa 13020 accacagtga ctctcagtg gaaccacg gaacaccag gggggatag accacaggggggaacca gctccagtg agaagagac caacttgga gagttggag tcaatgcaa 13200 acagaaacaca ccctagattg ggttgcctg ctttttaaag ccaactgtga gaagacattct 13320 acagaaacaca ccctagattg ggttgcctg ctttttaaag ccaactgtga gaagagacca 13320 acagaaacaa dgcttcag tacattctgg agaaagccc actagatag tgccacagg 13440 acagaaaaga agacttccag tacattcagg agaaagacca actagataga tctgggaga tcagtgaca actactacag ggtgacactac tttgatcaac ttacaaaaa aaacttgga atgacaatt tctgggaca tttgatcacac tacaaaaa aaacttgga agacactag tctgggaag tcagtgaca acacacttg aggacattcc tctgcaa aggaccca acacacttg acagataga agcacacag tacattcaag aggaccca aggacacac acacacttg acacacattg aggacacac aggacacac acacacttg acacacatta gaacactt tacacacaga gagacacaca aggacacaca cacacacttg acacacatta gaacactt tacacacacacacacacacacacacacacaca	ggt.gggagaa	ctgaagttca	gttagttcag	tatggcaaag	aaaaggcaaa	taaagacaga	
ggcaggaatt caatattga tgaaagatt agaaacata agccttcat acaattga 12840 atcoggaaca ggaattctat ctcaaaa tttogatta atcaaagag gaactgat 12840 atgaccatct totgttggt atttggtaga ttatggtagt cacacttott ccaaatattt 12960 ttacaggaca cactaccatt atcagcaca gcaatatga tcataccata 12960 ttacaggaca coctggaga tggtagct ccagctttac cacacaaca agctaagaaa 13020 aactgttgga accaaattca ttatttacat tttacacaag acceptagaaga tcaataccat 13080 gaaactgttga tgttctatct ttcttcaaaa aactgctcc taaaggagga atcatacatt 13080 gaaactgttga tgttctact ttctctaaaa aactgctcc taaaggagga tcaatacaa 13080 gagggagatatg atcaagaaga ggggggggggggggg	ctacttgcag	gatectcaag	taagccattg	acgtggaaat	taatagtttg	ggaagtagta	
accoggaaca ggaattecta ctcatcaaaa ttctgcattc atacaagagg gaacctgatt 12900 gcaaatcaga catcacctatt tctgttgdaga ttattgtggtt caacacttctt ccaaatatt 12900 gcaaatcaga catcaccata atacagacaa gctaagaaa 13020 ttacaggaca cccctgagaa tgggtagcct ccagctttac cacccaaacaa agctaagaaa 13020 aactgttgga accaaatca ttattacat tttcaacaag acctgaagaa tgttctact tctcttaaaa aatctgctcc taatggtgg tattctacatg 13140 ataatcgtgt tctaatccag tggaacctga cgaaaatgga aggsttggag tcaatgcaag 13200 gggggatatg atcagagaat gtctgtgatc gtgtcctgag aagcaccagg aacacctttg 13200 acctcaagtga ctctcgattg aagaagaac aagttgtat tgacagtgg ttgggaactat 13320 acagaacaca ccctagattg ggttytcctg ctttttaaag ccaactgtga gagacattct 13320 acagaacaca ccctagattg ggttytcctg ctttttaaag ccaactgtga gagacattct 13320 acagaacaca ccctagattg ggttytcctg ctttttaaag ccaactgtga gagacattct 13320 atagaccatg cgttsggag tcagtyatcc agctaaccagg gggttatcg crggaacatt 13320 atagaccatg cgttgggag tcagtyatcc agctaaccag aggatagt tcctggaacat ttgtacaac ttatcaaaaa aaacattgga atgacaattt ctggggcag 13500 atgaccatgt gctgggaga tcagtyatcc agctaacaga gggactatcg crggaacat 13200 atgaccatg gggggtac aacacacttg aactacaca gtgtaattc agcaacatg gtcaacaga agctcctca aggtaggac aacactggaagac caacagttga ttttttaa aggtaatgaa aactacaca gtgtaattc agcaacatg gtcaacaga 13740 aggtacaca aggtaccaca caacacttg accaacacttg aggtaggtca cacacacactg aggaagacca cacacacttg aggtaggtc tttaacacaga 13740 aggaaaaaa cccttggca ttttttata aggaagaca agaccaca cacacacttg aggaagacaca aggstggcca gaaaaattg aggaacaca cacacacttg aggaagacaca aggstggcca gaaaaattg aggaacaca cacacacttg atgacacaga tcaacacaca aggaagaca aggstggcca gaaaaattg aggaacaca cacacacttg aggaagacaca caccagaa ttcaacacaga tcaacacacaca aggaagacaca acaccaggaa agaccaca cacacacttgta ttcaacacacacacacacacacacacacacacacaca	ggcaggaatt	caatatctga	tgaaaagatt	agaaacataa	agccttccat	cacaattccc	
acqacatct tetgetggc atteggtaga tetatgtggtt cacactett ccaaatattt 12960 tcacaggaca caccacatt atcagcaca gctaatagca tcatecagga atcatacat 12960 tcacaggaca cccctggaga tgggtagcct ccagctttac cacccaaaca agctaagaaa 13020 aactgttgga accaaattca ttatttacat tttcacacaag acctggagaa tcatacatt 13080 gaaacgttga tgttetact ttcttaaaa aatctgctc taarggagga tcatacatg 13140 ataatcgtgt tctaatccag gtgaacctga cgaaaatgga aggtttggag tcatacaag 13260 gaggggatatg atcaqaagat gtctgdatc gtgtcctgag agacacctgga gacaccttg 13260 acctcagtga ctctcagtt gaggagacc cacgtgtat tgaccagtgg ttggggactt 13360 accgaaacac cccatgattg ggttgtcctg ctttttaaag ccaactgtga gagacttcg 13260 acgaacaca ccctagttg ggttgtcctg ctttttaaag ccaactgtga gagacttct 13380 gggggaacta tgcttcagt tctaccatg ctgcatatga tgtagtggag agacgttag 13440 aaaatgagac agactccag tacattctgg agaaagccc actagatagt gtccaccagg 13560 ctggaccaca tttgatcaac ttatcaaaaa aaaactgga atgactaatg ctgggagact caggacatg caggactacac ctggagactac atgactataga agactacaac ggggggtatac cacacacttg acggaacatt ttggagaga gagggggggggg	acccggaaca	ggaattccta	ctcatcaaaa	ttctgcattc	atacaagagg	gaacctgatt	
gcaaatcaga caccagaga tgggtagcct caagttttac acaccaaca agctaagaa 13020 aactgttgga accaaattca ttatttacat tttcaacaag atctggagaa 13020 aactgttgga accaaattca ttatttacat tttcaacaag atctggagat accaattcaat 13140 gaaacgttga tgttcatctc tctcttaaaa aatctgctc taatggtggt atctacatga 13140 gaaacgttga tgttcatctc tctcttaaaa aatctgctc taatggtggt atctacaatg 13140 accagagatga tacaagaga gggagattat gacaagaga ggggagataa accactagag ggggagataa accactagag ggggagataa tcaatacaag gggggagacaca tgtttcaatg tggttgtcctg ctttttaaag caagttgta tggacagag agaacacttg gggggagacaca tgtttcaag tctaccaagt cttcaaaaa aaaacatgga gagacattct 13380 gagagaacaa accactagatg ggggaacacaa tgaacacaga aaaatagaa gacactccag tacattctgg agaaagacca aaaacttga gggggaactac tgttgagagaa tcaatactag ggtgagacacaa tgaacacaga aaaactaga gagacattc 13380 ataacacaa tgtgacaaa ttaacaaaa aaaacttgga aggacattt ctgggacaat ttatcaaaaa aaaacatgag atgacaattt ctgggacaat ttatcaaaaa aaaacatgag atgacaattt ctgggacaa ttattttaa aggaatgaa aacacacta ggtaattca aggacacat gggacacaca gggagacaca aggagagacaca aaacactaca gggagacaca aggagagacaca aggagagacaca aggagagacaca aggagagacaca aggagagacacaca aggaagacaca aggagagacacaa caacacatagaga ttaacacaca gggagacacaaga ttaacacaca gggagacacaa acacacacagaga taccacaagaa tggagaacaca aggagagacacaca aggagagacacaca acacacac	atgaccatct	tctqttqgtc	atttggtaga	ttatgtggtt	cacacttctt	ccaaatattt	
tactactagga cacaatta tattacat tittaaataa attiggaaga tatatataat 13080 gaaacgttga tgttctatct tctctaaaa aatctgctc taatggtggt attataat 13080 gaaacgttga tgttctatct tctcttaaaa aatctgctcc taatggtggt attataaca 13400 gaaacgttga tgttctatct tctcttaaaa aatctgctcc taatggtggt attatacaatg 13400 acatcaggggagatatg atcagagat gtcgtgat gtgtctctg gtgtcctga gagacacagg aacacctttg 13260 acctcagtga ctctcgattg agagagaaga caagttgtat tgatcagtgg ttgggagacaca tcctagattg gtgttcctg cttttaaaa caagtagg agagtctct 13380 ggggaactca tgctctcagt tctaccatg ctgcatagg gagacattct 13380 ggggaactca tgctctcagt tctaccatg ctgcataggagggagacacacacacttggagaggaga	gcaaatcaga	catcaccatt	atcagcacaa	gctaatagca	tcattctgga	atcatcacta	
gaaacgttga tgttctatct tctcttaaaa aatctgctc taatgggg attcaacag 13200 gagaggtatag atcaagaat tgttgtgatc gtgtcctgag aagaccagg aacaccttg 13260 acctcagtga ctctcgattg aagagaaga caagttgtat tgagtcagg aacaccttg 13260 accgagacaca cccatgattg ggttgtcctg cttttaaag ccaactgg gagagacttc 13320 acagaacaca tgcttctagt tctaactcatg ctgcatatga tgtaggggag gaagtgctag 13440 aaaatgagac agacttccag tctaccatg ctgcatatga tgtaggggag agaggctag 13440 aaaatgagac agacttccag tacattctgg agaaagccc actagatag tgtcaccagg 13500 ctggacacaa tttgatcaac ttatcaaaaa aaaacttgga agacaatt ctgggaccat 13500 ctggacacaa tttgatcaac ttatcaaaaa aaaacttgga agacaatt ctgggccag 13620 attaccttag agggggttac cacacacttg accagtagat ctcttggacaacat 13200 ctggacacaa tttgatcaac aacacacta gtgtaatcc agcatcataa gtcagacagt ttattttaa aggtaatgaa aactacacca gtgtaatcc agcatcataa gtcagaacag agggcctcc atgttttgtaa tgttgacca cacagttgaa tgttgcagaga ttaaatggc ccaatattgg ccagaacca cacgagagtc accagtagag tgctaccaag cctagaagaa ggagaccatg tctttactt caggaagata cactttgcc ttaaccaga aggacacatg ttcaggaact tcttactt caggaagaca acactggga tcaagagaaa acccttgtgt ttcttgaag tccagtacag agggtggcct gaccaaga 13980 ccaatattgg ccagaacca tctttactt caggaagac agggtggcct gaccaaga 13980 ccaatattgg cagaacactg tctttactt caggaagac agggtggcct gaccaaga 13980 ccaatattgg agagacactg tctttactt caggaagaca gaggtggcct gaccaaga 13980 ccaatattgg acagacaca tctttttctgaa tgttttcat atgaaaaca acaggggcada 14040 accctgatga ttcagtgact tctttactga tctcattga tgaagacaca gggttctaa acaggaaaaa ccccttyttg ttcttgaa gtgttgcaga agaggggct gacaagagga 14040 acctgacaga gaaaaaa cacttgttg ttcttcatga tgaagaga gaggggct ggaagacaca gggttctaa acagaaaaaa caccttgttg ttcttcatga tgaagaga gaggggct ggaagagaca 14100 agacaaaagga acaacaa tgaagagaca tggaggggg tggaagaca acactgtca 14280 agcaaaagga acaacaa tggaagaga gtgaagaga gaggggga gaggagaga 14460 tactatgaa acaggaa tactattat agaaggtta acattatt gactttaga acactgtta 14800 agcaaaaagga acaacaa tctatcaata acaggagaa acaggggaa gacagggaa 14460 tactataaaa gtcaaggaa tcttttacaa aagggttta tattttattt	ttacaggaca	cccctggaga	tgggtagcct	ccagctttac	cacccaaaca	agctaagaaa	
ataatcgtgt tctaatccga gtgaaccta cgaaaatgga aggttiggag tcaatgacagggggggggggggggggggggggggg	aactgttgga	accaaattca	ttatttacat	tttcaacaag	atotggaaga	tcatattaat	
gggggatatg atcagaagat gtctgtgate gtgtcctgag aagcaccagg aacacctttg acctcagtga ctctcgattg aggaagaga caagttgtat tgatcagtgg ttgggacttt gggggaactca tgcttctagt bctacctatg ctgcatatga tgtagtggag gaagtgctag ggggaactca tgcttctagt bctacctatg ctgcatatga tgtagtggag gaagtgctag ataactgtgg gctggggag tcagtgatca ggcacacgag gggcttatcg ctggaacat ctggacacaa tttgatcaac ttatcaaaaa aaaacttgga atgacaattc tggggagat cagttatcaa gggacattct ctggacacaa tttgatcaac ttatcaaaaa aaaacttgga atgacaatt ctggtgcag attaccttag aacctttgaa aactacatca gtgtaattcc agcaacatt ctggtgcag attaccttaa aggtatgaa aactacatca gtgtaattcc agcacactaa gtcagacaa tgcttgtaaa ggggggttac cacacacttg aacagatttt tygcagatga ctgggaaca aggctcctc atgttgtaa tgftgaccac accagatttt tygcagatga ctgggaaca aggctcctc atgttgtaa tgftgaccac accagatttt tygcagatga ctgggaaca aggacaaaa cccatgtgaagagagac tccagtaaa tggtgaagg ttaaactgac caggaagaaa cacagagagtc atccataaa gggaagagtc taaccaaag aggagaaaaca coccttgtgtg ttcttgacacg accatgctgt ttaaccaaga aacctggaagaaaaca coccttgttg ttcttgag tgctgaagagaaaaa agagagaaaaa cccttgttg ttcttgag tgctgagagt aggagaacaca gggttctaa aacaggaaaaa cccttgtgtg ttcttgaag tgcagaagaca gggagacaca acctggaagaaacac coccttgttg ttacaggttt gcgtgagag tacagtaaa tgcgacaaaa aggagaaaac cccttgttg ttacaggttt gcgtgagag tacagtaaaaaagac accttggaagaaaaaagac tttgtcaattg actcattagaa tggcaaaaaaagac accttggaagaaaaaagac tttgtcatcg gtaagaacaaa tgagagagagaa gtgagacagac acctagacaat agaaaaagaa attatgagaa caaggtttaa accagttt gaaagaacaaa agaaaaagaac tttgtcaattg aacctaattgaaa accagagaaaaaagac tttgtcatcg gaaaacatt ataaaacat tgaaaataa acctttacaaaa gtcaagaaa accacacttaaaaaagaac accttttagaa gaaaacagat accatttaaaaa gtcaagaaaaagaac accacaattaaaaaagaac accacacacaagaaaaagaac accacacaca	gaaacgttga	tgttctatct	tctcttaaaa	aatctgctcc	taatggtggt	attctacatg	
gggggatatg atcagaagat gtctgtgatc gtgtcctgag aagcaccagg aacacctttg acctcagtga ctctcagttg aagagaagac caagttgtat tgatcagtg ttgggactt acagaacaca cccatgattg ggttgtcctg ctttttaaag caactgtga gagacgatct gggggaactca tgcttctagt tctaccttatg ctgcatatga gtgatggaa gaagtgtag aaaatggac agactccag tcaattctag agcaaatcgag agaagtgctag atgaccatg gctgtgggag tcagtgatcc agcaacacg gggcttatcg ctggacacaa tttgatcaac ttacaacaaa aaaacttgga atgacaatt ctgggacaca tttgtttaa aggtaatgaa aactacatca gtgtaattcc agcatataa gtcagaacac tgcttgtcaa ggggggttac cacaccttg aacagattt tcttatattga gtcagaacaca tgcttgtcaa ggggggttac cacaccttg aacagattt tggtagaaga gtaagaacac ccaatattgg ccagaaccac aggaagtc atcattatga tgttggacac acaagttgaa tggtagaaga gtaagaacac ccaatattgg ccagaaccac acaggaagtt atcatataga tggtagaaga ggaagacacg ctgagaagaa ggaagcactg tctttatctt caggaagata tggtaccaag ccttctgcca gaaaaattag agagtcatca atcacgaaga tccagtacag agggtggct gacaagacag gaaaaaaa cccttgttg ttcttgtgat tgtttttaa atgcaaaata agagggcag acaagaaaaa cccttgttg ttcttgtgat tgtttttaa atgcaaaaaa acagcatgg gtagagaaca tctttttgaa acaggaaagaa aaaaaagaac tttgttgaat tgttgagaca agagtgctgg acaagaagag aagaccatg tcttatactt gaaggacggg ggaggtggt gacaagaga gaagaacac tcaggaag tcaggaagacagggaagaagagaag	ataatcgtgt	tctaatccga	gtgaacctga	cgaaaatgga	aggtitggag	tcaatgcaaa	
acagaacaca cccatgattg ggttgtcctg ctttttaaag ccaactgtga gagacatct 13380 ggggaactca tgcttctagt tctacctatg ctgcatatga tgtagtggaa gaagtgctag 13400 aaaatgagaa agacttccag tacattctga agaaagccc acagatagt gtcagtggaa 13560 atgaccatgt gctgtgggag tcagtgatca agctaaccag gggcttatcg ctggaacat 13560 atgaccatgt gctgtgggag tcagtgatca agctaaccag gggcttatcg ctggaacat 13560 attactttta aagctatgaa acaacactag agatgattt tcttatgat gttacatgg ttattttta aggtaatgaa aactacatca gtgtaattca agcatatta ctgggaaca 13620 ttattttaa aggtaatgaa aactacatca gtgtaattca agcatata gtcagaacag 13740 aggcttgtoaa ggggggta cacacacttg aacagttta tygcagaga ttaaatgacc 13860 ttattttaa aggtaatgaa acacacatca acagttgaa tgctacaaag ccttctggcaa aggagacca cacacacttg aacagttta tygcagaga ttaaatgacc 13860 ttattttaa agggaacca cacacacttg aacagttagaa tgctaccaag ccttctggcaccaatattgg ccagaacca acagaagtta atcatatgga tgctaccaag ccttctgcca 13920 acadaattgg agagtcatca atcacgcaga tccagtacag aggggggct gacaaggagaaaattag acagcatgt ttcttgtgat ttgttttcat atgcaaaaata agaggggctag 14040 accctgatga ttcagtgaat ttctttgggatt tgtttttcat atgcaaaacaa agagggatag 14100 acctgatgaa ttcagtgaac ttctttgtgat tctcattgaa tgcagaaacca gcgttcttaa 14160 ttacataggaa acaagcaatg gtctcattga tctcattgaa tgcagacaa cacctagcaag 14200 accaggaaaaaacaa gagaagaca tttgtcaattg gtgagacaaca gagaagaca tttgtcaattg acagtaattt gcgtgaaga aaaaaagaac tttgtcaattg gttaggtcc aatttggacaa acacctattacaa gaaaggttt tgcttattt agaagagaca 14200 gcttttata attatggga attataacatt taaatggta acacgttt aacaggaa agaaactt taaacgtaa tataaacgat tacactatta acacttaac ggaagagat agacagaca tttacactaa agaagacttt aacacgaa agaaactt taaatggta tacaaagac aggagagat agacagaca tttaaacagaa agaaactaca tcttctgtga atcaaatga tattgaacagat agaacagaca gagaattta acactatta tagtggggtt tacaaagaga agaacagaca tcttctaacaa agaagacaa ggagacagaca tctttaacacaagaa agaaactaca tcttaatgga agaacacagaga tacacaagaa agaaacacaa caccaagagaa aacacacac	gggggatatg	atcagaagat	gtctgtgatc	gtgtcctgag	aagcaccagg	aacacctttg	
agagaactca tgcttctagt tctactag cgcaataga tgtagtggaa gaagtgctag aaatgagaactag tgctgtgggag tcagtgtgg agaaagccc actagtagt gccacaagg ttacattctgg agaaagccc actagtagt gccacaagg ttagtagacaatg gctgtgggag tcagtgadc agctaacag gggctatag ctggaacatt ttgatcaac ttatcaaaa aaaacttgga atgacaattt ctggaacatt ttatttta aggtaatga caattaagtag ggagtattc cacacactg aaaatagata gagatattt tcttatagt gtcagaacag ttattttaa aggtaatga aactacata gtgtaattc agctaatag tcagaacag ttattttta aggtaatga actacatag gtgatttc aggtaattt tgycagatga cttggaacaa aggctccc atgttgtaa tgttgacaa acaagttga ttgtgacaag ttaaatgac caaggaagtt atcattatgg cgaaaaatag gagagcact gttttactt caggaagata tctttactga gggaagaaca ggaagacad atcacactt caggaagata cacactgctg ttaaccaaga 13980 ccagaagaag ggaagacad acaccacttg acaagttgaa tgtgacacaag ggaagaacaa acacgaagaagaa acacgagaagaa acacgagaagaa ccccattgaga tcaattgaga ttctttgcag tgcttggaag acacgagaagaa acacgagaagaa ggaagacad ggctcttaga tgttgtagaag tcattga tgttgtaaa tgcagaacaa gggggacaa acacgagaagaa ccccttgtg tttctggaag tcattga tgctggaagt ggaagaacaa gggtgttt gcttataga tgcagaacaa gggtgttaa acacgagagaa tactatttg aaaggtggct gaagaagaca gggttttaa acaaggagaa tactatttg aaaggcagga ggaagaaca ggttgtagaag tactatttg aaaggtgtt acattttggac ggaaagaaca ttggaagaaga ggaagacat gttgtatta acaaggggaa agaagacat ttgtaattg agaacgat ttgtaattg agaagagaa gaaaaaagaa gaaagaacat ttggaaattt tggtagaagaa agaagagaa gaaagaaca ttggaagaga ggaagacag ggaagaagaa acacgagaa acacgagaa tactatttagaagaagaa ggaagaaca agaagagaa gaaaaagaac ttggaaattt aacaagggga agaagaaca agaagagaa gaaagaaca ttggaaatta acacgtatt aacaagggga gaaaaagaa acacgagaa acacgagaa tacaattt taggtggaa aaaaaagaa gaaagagaa gaaagaaca ttaaagaagaa gaaaagaa atacacgagaa acacgagaa acacgaga	acctcagtga	ctctcgattg	aagagaagac	caagttgtat	tgatcagtgg	ttgggacttt	
aggadactat tgettetag tacattetag agaaagace gagaagat gaaaatagaga agacttecag tacattetag agaaaagace acattagatag gtecaceagg atgaccatg getgtggag tacagtgate agctaaceag gggettateg etgacatat teagtgacatattettata agctatted aaaatagata gagatagtt teettatga ttacattaga attacettag aggettatee aaaatagata gagatagtt teettatga tetggeaaa tattetttaa agggaggtac cacacactt gagaagatt teettatga tetggaacat ggggettatee atgtttgaa acatacatea gtgtaattee tetggaacat gggettatee atgtttgaa tgttgaccae acaagattt tygcagatga ttaaatagac etagagagate etatatet caggaagate tgtgacaaga tutaaatagac etagagagate atcetatgga tgggagagagagagaagaagaagaagaagaagaagaagaa	acagaacaca	cccatgattg	ggttgtcctg	ctttttaaag	ccaactgtga	gagacattct	
atgaccatgt gctgtgggag tcagtgatca agctaaccga gggcttateg ctggtaccat ctgagcacat ttatacaaaa aaaacttgga atgacattt ctggtgccag attactttag aagctattga aactactca gtgtaattca agcatcataa gtcagaacag ttatttttaa agggcatca cacacacta gtgtaattca agcatcataa gtcagaacag tgcttgtcaa ggggctccc atgtttgtaa tgttgacca cacacactta atccatgag tgtgaccaa acactgcggt ttaacacaga aggagaagata atcctatgga tgtgacacaa acactgcggt ttaaccaaga aggagaagata atccatgaga tccagaagata tccaggaaga tcaggagaagata accatgagag tccaggaagataaaattag agagcactg tctttactt caggaagata acactgcggt ttaaccaaga acactgagag acactgaga acactgaga ttcaggaaga cccttggggt ttcagagaacaa acacctgaga acactgagag acactgggat ttcagtgat ttcttggatt tgttttca atgcaaaata agagggcag acactggggt gacaagagacag cactggagacacagaacactagagaacactggagaacactggggt gacaaaagacagacaggaaacactggggt gttcattga acacgacaga acactggggt ttaccatgaa acagcaaga gttcaatga tccattgaa tgcagaacaa agggtggcct gacactgggg acactgggaacacctggagacacctgagaacacggaaacacctggggt gacaaaagacacagggtagacacaggaacacgggtagacacaggaacacggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacggggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacaggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacaggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacaggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacaggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacaggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacgggaacacaggaacacgggaacacgggaacacgggaacacagggaacacgggaacacgggaacacagggaacacaggaacacgggaacacgggaacacacacacacacacacacacacacacacacacacaca	ggggaactca	tgcttctagt	tctacctatg	ctgcatatga	tgtagtggaa	gaagtgctag	
atgaccatgt gttgggggg teatca space sp	aaaatgagac	agacttccag	tacattctgg	agaaagcccc	actagatagt	gtccaccagg	
ttattittaa aggtaatgaa aactattaatat tittittaa aggtaatgat teetitagat gteaacaaga tittittittaa aggtaatgaa aactacatca gtgtaattee agcateataa gteagaacaa tigettytea ggggggtae cacactig aacagtitt tyycagatga ettigggaaca taggteetee acaacactig aacagtitta tyycagatga ettigggaaca taggeeteetee acaagtitgaa tgtggeagag ttaaactagae eaggaatteetee eaggaagate ateetatgaa tgtggaagate ateetatgaa tgtggaagate ateetatgaa tgtggaagate acactgeetit taacacaaga agggagatea acactgeetit taacacaaga acactgatgaa teetagaagaa acactgeetit taacacaaga acactgatgaa teetagaaaaa ateetagaaa acactgatgaa teetagaaaaa acactgatgaa teetagaaaaa aggggagaa teetatteetagaa tgtaagaaaaa agggggaagaa acactgatgaa teetagaaaaa teetagaaaa teetagaaaa teetagaaaaa teetagaaaaaaa teetagaaaaaaaa teetagaaaaaaaaaa	atgaccatgt	gctgtgggag	tcagtgatcc	agctaaccga	gggcttatcg	ctggaacatt	
ttatttttaa aggtaatgaa aactacatca gtgtaattcc agcatcataa gtcagaacag 13800 cocatatttttgaa tgttgacaa cacacttt 15956agatga ttaaatggac 13800 cocatatttgg cagaaccac caggaagttc atcetatgga tgtgtacaag tgtgtgaagag ttaaatggac 13920 ctgagaagaa aggatcatca atcatcaggagat cacactatgga tgtgtacaag ggagagaatattag ggaagacatg tctttatct caggaagatc acactgctgt ttaaacagaagaacactggagaaaaattag agagtcatca atcatggaat tcaggaagaat cacatgagga taacattaggaa acactgaggat tccaggaagaa cacatgaggat taaggagaaaaattag agaggaaaaa tgagaagaga tgtgtagaagaagaacaaggaagaacaaggaagaagaacaaggaagaa	ctggacacaa	tttgatcaac	ttatcaaaaa	aaaacttgga	atgacaattt	etggtgeeag	
tgettgteaa ggggegttac cacacttg aacagattt tggcagtag cttgggaca 13800 aggeteete atgttgtaa tgttgaccae acaagttgaa tgtggcagag ttaaataga 13800 ccaatattgg ccagaaccae caggaagtte acacattgg tgetaccaag cettetgaca 13920 ctgagaagaag ggaagcactg tetttatett caggaagate accettgtg ttaaccaag 13980 gaaaaattag agagteatea atcacgcaga tecaggaagate accettgtg ttaaccaaga 13980 accettgatga ttecagtgact tetttatett caggaagate accettgtg ttaaccaaga 14040 accettgagat tetttatett gettgagatt ggaagaaca gegttettaa 14100 accettgagaa acaggaagaagaagaagaagaagaagaagaagaagaagaa	attaccttag	aacctttgca	aaaatagata	gagatagttt	teettatgat	gttacatggc	
aggetcotec atgettigtaa tgitgaccac acaagittgaa tgitgacagag tiaacagagagagagagagagagagagagagagagagagag	ttatttttaa	aggtaatgaa	aactacatca	gtgtaattcc	agcattataa	gttagaacag	
ccaatattgg ccagaacca caggaagttc atcatagga tgctaccaag ccttctgca ctgagaagaa ggaagaccg ctcttatct caggaagatc acactgctgt ttaaccaaga 13980 accctgatga gaagacactg tctttatct caggaagatc acactgctgt ttaaccaaga 14040 accctgatga ttcagtgact tctttatct tgtttttat atgaagaacaa cccttagtag ttccagtagat tcttttatct tgtttttat atgaagaacaa acccttagtaga ttccagtagat tcttttatcat atgaagaacaa accctagtagag 14100 accaggaaaaaa cccttgttg tttcttgcag tgctggagtt ggaagaacaa gggtggcag accatggag 14100 accaggaaaaaa cccttggtagagaaaaa tttactggaa acaggcatgg tttctattgaa tgcaggaggac accattggaa 14160 accaggaaaaaaaagaa accaggagaagaca gtgaggcggg atgaggacaaa atgacaagaa acactagtca 14280 attactagaa gaggagagaa acaacggggaa acaacggggaa acaacggggaa acaaagaggc ttggtgaagaa 14340 acatttaaaaa gttcaaggaa ataaaaggttt acatattta gaaaaggggaa gagaagagaa	tgcttgtcaa	ggggcgttac	cacacacttg	aacagatttt	tataaaaaa	ttaaatgacc	
ctggaagaa ggaagcactg chitatett caggaagate acactgctgt ttaaccaaga gaagacatg gaaaaattag agagtcatca atcacgcaga tccttatett caggaagate acactgctgt ttaaccaaga accctgatga ttcaggatt tcttgattt tgttttcat atgcaaaata agagggctag 14040 tectaggaaa acaggcatg tttettgaa tgcaggaagate ggaagaacaa ggattatte ttatacaggaaaaa cccttgtgt tttettgaag tgcaggagggggggggggggggggggggggggg	aggctcctcc	atgtttgtaa	tgttgaccac	acaagutgaa	tactaccaaa	ccttctgcca	
gaaaaattag agagtcatca atcacgcaga tccagtacag agggtggcct gaccatggag 14040 accctgatga ttcagtgact ttcttggatt tgttttcat gagagaaaata agagggctag 14100 accaggaaaaa ccccttgttg tttcttgcag tgttggagtt ggaagaacca ggattcatta 14160 accaggaaaaa ccccttgttg tttcttgcag tgttggagtt ggaagaacca ggagggctag 14100 accaggaaaaa ccccttgttg tttcttgcag tgttggagt ggaagaaca aggaggctag 14100 accaggaaaaaa caccttgtt gttctgatgt ggaagaacaa aggaggctag 14100 accaggaaaaaaaaaaaaaaaaaaa ttgaaaagaa ttgaagaagaa gtaggagaga acacattgaa aaggaggctag aagaaaaaaaaaa	ccaatattgg	ccagaaccca	caggaagttc	accctatgga	acactactat	ttaaccaaga	
qaadaattag agagtaatta teetagaat tytettetaa agagaaataa agagggetag teetatgaa teetagaata tyeetagagt tyeetaataa agagaaaaaa cacctagtaa teetagaagaa teetagagagaa teetagagagaa teetagagagaa teetagagagaa tyeetaataa tyeetaataa tyeetagagata agagagaaaa agagagata tacaagtata tacaagtata tyeetaataa tyeetaataa aagaagget tyeetaataa aagaaaggaa tacaagtata aagaaggaa teetatte aaaagataa tyeetaataa tyeetaataa aagaaggaa tyeetaataa aagaaggaa tyeetaataa aagaaggaa tacaatatta aaaagaaa teetagaaaaaa tacaagtaa tyeetaataa aagaaggaa tacaagtaa teetaagaa tyeetaataa aagaaggaa tyeetaataa aagaaggaa tacaagtaa teetaagaa aagaaggaa tyeetaataa aagaaggaa tyeetaaaaaa tyeetaataa aagaaggaa tyeetaaaaaa tyeetaaaaaa tacaagaa tacaagaa tacaagaa tacaagtaa aagaaagaa tacaagaa aagaacaaa aagaaagaa tacaagaa tacaagaaaaaaaaaa	ctgagaagaa	ggaagcactg	tetttatett	taggaagate	acactgetge	daccataga	
caaggaaaaa cocttgttg tttctgcag tgctgagtt ggaagaacca gcgttctaa tactatggaa acagcatgt gtotcattga tctcattgaa tgcagtagc ggaagaacca ggaaaaaac tttgtcattg ggaagaacca ttggtcaaca ggaagaggaa acaacagggta tactatttg acagcattgt ggaagaacca ttggtcaaca ggttggtcaaa acacctagtca ggttggtaaga acaacagggta gttagtattg ggaagaacca ttggtaaatt ggaagaggaa gaaaaaaca ttggtcaaca ttggtaatt ggaagaggaa gaaaaaaca ttataaacat gaaacggtt taaaacat tattaggaaca acactttggaaata tatttggaagaa ggaagaacaa ggaggagaa acaacaggggaa ggaagaacaa gaggggaagaa gaaataggaa ggaagaacaa tattagaaca acacttatca ggaaacggaa ttaaaacaat ggaaacgaacaacaacaacaacaacaacaacaacaacaac	gaaaaattag	agagtcatca	atcacgcaga	tattttcat	atocasasta	agagggctag	
tactatggaa acagccatgt gtctcattga tctcattgaa tgcagtcagc cagtttattc actagcatg gtaagaacaa tgagagagca gtgagccgtg atggtccaaa cacctagtca 14280 1420 acagcaaagga aaaaagaac tttgtcatct gttaggttcc attattgca tgataattgt gttaggttcattg attattgggc aagtagctgt ttgctatttt gatcattatt actattcaat gaaacgtttt actattcatt gatcatcttt tacagggagggggggggg	accetgatga	ttcagtgact	ttetggattt	tactageatt	acgedadacca	gcgttcttaa	
ttacagtttt gegtgtgaag aaaagaaca tgagagagaa gtgagcegtg atggtccaaa Cacctagtca 14280 aagaaaaagaac tttgtaattg attattggg aagatattt actattcaat gaaacgttt aaaaggggta gtaaaaagac tttgtatattg aattattggac taatattta acattcaaa gttcaagtaa ttataaaaat tgaaattct tatacagtga ataacctaat tataaaaat tatattcgtcg gaaaaagaact gtaaacgat ttaaaaaat tatattcgtcg gaaaaagact agaaattct aaaagaat tataaaaaat tatattcgtcg gaaaaaagact agaaattct aatagctga ataacctaat ggaaattct tataccatta accttaac aggatttta acaatgatta ataaaaaagt ccacttaac ggaaaaagact gtaaacggggaa gaaaaagact agactagcta ataaaaaagt tatacgaaaa aggagagaa agactagcta ggagagagaa aggaggagagaa agactagcta aggaggagagaa agactagcta aggagagagagagagagagagagagagagagagagag	caaggaaaaa	ccccttgttg	atataattaa	tctcattgaa	tacaatcaac	cagtttattc	14220
ttacagtttt gcgtgtaaa tactattttg aaagcttatt gagaaggctt tggtgaaga 14400 gtttgtattg attattggc aagtagctgt ttgctattt gatcttattca tgataattgt 14400 gtttgtattg attattggc aagtagctgt ttgctattt gatcttattt cagaagggca 14460 tactattaaa gttcaatgta aagtagctgt ttgctattt gatcttattt cagaagggca 14460 tgctttacag cagaaatctt ataatgata actggtaata tatttcgt: gcaaaaaat 14580 tgctttacag cagaaatctt ataaacat tgtaaattgt ataatgtaat gggagagaga agactagct ggaaaattctt tatagctgta cttctgtgta atcacaagact ggggagagaat ggactaggt ggaaaagatttga agactagct atacctatt ggggaagaat aatgacagat taacctaatt tagttggta ataaaaagtt ccatttcatg gggaaaagagtggaagaatgagagagagagagagag	tactatggaa	acagecatge	tragararca	ataaaccata	atggtccaaa	cacctagtca	14280
agcaaaagga aaaaaggac tttgtcatct gttaggttcc attattgca tgataattgt 14460 taataatttt actatacaat gaaacgttt aaacggggta gaaaaagact agtttttga ttgcttacag cagaaatct ataaagact ttataaaaa ttgtaaatt tattgctg gcataaaaat tattgaaat tattgaaat tattgaaat ggggaaagactaggaaattaaacaat tgtaaattgt ataaaaagact ggggagagaa agactagcta taacattaaaa gttcaaagga atcaactaac agagttttga ataaaaaggt caattgaaat ggggaagagaa agactagcta gggaaagaatggaaatgaggaaatgaggaataaaaagactaggggaagaatgaggaaatgaggaaatgaggaaatgaggaaatgaggaaatgggaagaa	ttagacatg	gcaagaacaa	tactattttq	aaagcttatg	aagaaggett	tgctgaagaa	14340
gattgattg attattgggc aagtagctgt ttgctattt gatcttatt cagaaggga 14520 taataatttt actattcaat gaaacgtttt aaacggggta gaaaaagact agtttttga 14520 tgctttacag gtccaagtaa ttataaacat tgtaaattgt attattggag gcataaaaat tgaaatctt taagctgta atcaatgatta actggtaata atattgaatc gcatttcat ggtcttctc ttatccatta accacttaac agagttttga ataaaaagtt ccatttcatg gggaataagaat aatgacaggt taacctattt tagttggta ataaaaaagtt ccatttcatg gagataagaat aatgacaggt taacctattt tagttggtta ctatgttcta ggtgttgtat gagagtagtt accactatt tagttggtta ctatgttcta ggggaggaggaggaggaggaggaggaggaggaggagga	accasacca	aaaaaadaac	tttgtcatct	gttaggttcc	atttattgca	tgataattgt	14400
taataattt actattcaat gaaacgtttt aaacggggta gaaaaagact agtttttgta 14580 14580 acatttaaaa gttcaagtaa tataacat tgtaaattg actattgaaat tatttagaat 14640 14680 ggataagaat aataccattaac agagtttga ataagaagat agactagcta 14700 14580 14680 14640 1	aycaaaayya	attattgggc	aagtagctgt	ttgctatttt	gatcttattt	cagaagggca	14460
tgctttacag cagaaatctt ataatgatta actggtaata tatttegtig gcataaaaat 14580 taaattaaaa gttcaagtaa ttataaacat tgtaaattgt atattgaaat 14640 tataaatctt tatacattaa ataactattaa acaaagact ggggagagat agactagct 14760 ggataagaat acaactatta ataactattaa ataactattaa agagttttga ataaaaagtt ccattcatg gggataagaat acaactatta tagttggta ataaaaaggt ccattcatg 14760 ggataagaat acaactatt tagttggta ataaaaaggt ggtgtgtaat 14820 gaagactagt ttacactat ttacaggca aagaaatagg tttggagggg ttgggagggaga ggagggggggg	taataattt	actattcaat	gaaacgtttt	aaacqqqqta	gaaaaagact	agtttttgta	
tgaaattett tatagetgta ettetetgtta ateaaagaet ggggagagat agaetageta 14700 getettete ttatecatta ateaettaac agagttttga ataaaaagte ceattteatg ggataagaat aatgacaggt taacetatte tagttggta etteteteta ggagtagtte acatagtte ateagaatete tettacaggea atgaaatagg ttacaatee tettagggg tettaggggggagaga gttactatta 14880 tetaceteat tettacaggea atteaggggggggggggggggggggggggggggg	tactttacac	cagaaatctt	ataatgatta	actggtaata	tatttcgttg	gcataaaaat	14580
gataagaat aatgacaggt taacctatt tagttggtta ataaaaaggtt ccatttcatg gataagaat aatgacaggt taacctatt tagttggtta ctatgttcta ggtgttgtat taccactcat ttacagtca aatgacaggt ttacaatccc aagagtttta actgattca attaagggggggggg	acatttaaaa	gttcaagtaa	ttataaacat	tgtaaattgt	atatgtaatc	atattgaaat	14640
getetttete ttatecatta ateaettaac agagttttga ataaaaagtt ccattteatg gataagaat aatgacaggt taacetatt tagttggtta ctatgtteta ggtgttgtat 14820 agagtagttt acatagtte actgattea ettacateee aggagggggggggggggggggggggggggggggg	tgaaattctt	tatagetgta	cttctqtqta	atcaaagact	ggggagagat	agactagcta	14700
ggataagaat aatgacaggt taacctatt tagttggtta ctatgttcta ggtgttgtat 14820 agagtagttt acatagttc actgatttca ctacaatccc aggagggggggggg	gatattata	: ttatccatta	atcacttaac	agagttttga	. ataaaaagtt	ccatttcatg	
ttacactcat tttacaggca aagaaatagg tttggagggg ttgggtgtt tgcccaagtt 14940 cccttgtagg gtagctaggg tcttatgtta cagtataagg gagagcagac tgttccttta aagagactga tcttatgtta cagtataagg gagagcagac tgttccttta ttattttat tgacgttagt attttacat gcacagggaa ggagggttt tgtttatttat tgacgttagt attttacat gcacagggaa ggagggttt ttttatttta	ggataagaat	: aatgacaggt	taacctattt	tagttggtta	. ctatgttcta	ggtgttgtat	
ttacactcat tttacaggca aagaaatagg tttggagggg ttgggtgttt tgccaagtt 15000 agaacctacc tcttagtggc tcttatgtta cagtataagg gagagcagac tgttccttta cccttgtagg gtagctaggg cttgtgaatt aagagactga ttaacaggag aagaggcata cacattttat tgacgttagt attttacat gcacagggaa ggagggttt attttattt ttattttat ctttattta aagagacagg ggtcttgctg tgttgccagg gctggactca attgtgaaga agtgactagt tcttctgctt gagattcctg agtagcaggg actataggg ttgtgagaat tgttctagtga ttttaacaa aggctgataa 15360 tgctcctctg tgcttggcta aagaaggggt ttgttgcagg gctggactca 25240 attgtgaaga agtgactagt caaagggaga qaggatttca gctccaaggg gtggtaaatt 15360 attgtgaaga gactaggaa tgtatagtaa tagtagtga ttgttaacaa aggctgataa 15360 attgtggtgt ggcatattct tctagctag gcaggtttgc ggaagggttt tgttaacaa aggctgataa 15420 ctgtatctgc tctaataag ggcaggaa acacctttac agatggaaat tcatatcacc 15540 ttccacagg gaaatttatg tcctgcctta ggcagttagg ggaagggcag agaattctt 15600 ctgtatctgc tgtgtccag gtgcttcag ctcaaaataa tccttatgcc aaagtagcat 15720 aaaggttttt tttaaaaaat catgtttcaa agatttgcatg tggaagacaa atggacatga 15780	gaagtagttt	acatagtttc	actgatttca	ctacaatccc	aggaggagta	. gttactatta	
ctcatcgtaa aatgacagat gaggattcaa attcaagtct taattgaagt ccattacttt 15000 agaacctacc tcttagtggc tcttatgtta cagtataagg gagagcagac tgttccttta 15060 cccttgtagg gtagctaggg cttgtgaatt aagagactga ttaacaggag aagaggcata 15120 cacattttat tgacgttagt attttacat gcacagggaa ggagggtttt attttattt 15180 aactcctgaa gccaagcgat tcttctgctt gagattcctg agtagcaggg gctggactca 15240 actcgtctg tgcttggcta aagaaggggt ttgtaggaaga gagggggggggg	ttacactcat	tttacaggca	aagaaatagg	tttggagggg	ttgggtgttt	. tgcccaagtt	
agaacctacc tettagtggc tettatgtta cagtataagg gagagcagac tgtteettta 15060 ceettgtagg gtagetaggg ettgtgaatt aagagactga traacaggag aagaggeata 15120 cacattrat tgaegttagt attracat geacagggaa ggagggttt attractt 15180 trattract ettrattra aagagacagg ggtettgetg tgttgecagg getggaetea 15240 aacteetgaa geeaagegat tettetgett gagatteetg agtageaggg actataggg attgtgaaga agtgaetagt caaaggagaa gaggattra geteecaggg gtggaaatr 15360 attgtgaaga ggaagggaa tgaaggggt ttgtaagga ggagggggggg	ctcatcgtaa	a aatgacagat	gaggattcaa	attcaagtct	: taattgaagt	ccattacttt	
cocttgtagg gtagctaggg cttgtgaatt aagagactga ttaacaggag aagaggcata 15120 cacattttat tgacgttagt attttacat gcacagggaa ggagggttt atttttattt 15180 ttattttat ctttattta aagagacagg ggtcttgctg tgttgccagg gctggactca 15240 aactectgaa gccaagcgat tcttctgctt gagattcctg agtagcaggg actataggtg 15300 tgctcctctg tgcttggcta aagaaggggt ttgtatgtga ttttaacaa aggggtaaatt gtggggaagat gactaggaa tgtatagtaa taaggtttgc tatttggca gtgggaaatt tcttccacagg gaaatttatg ggacagggaa acacctttac agatggaaat tcatatcacc tgtatctgc tgtgtccag gtgccttcag ggagttgc ggaaggggaa agaagtggca agaagttttt tctaacaag ggaaggttt tctaacaacagca tcttatgggtg ggcatatct tcaacagcat catctatact taacaacagca attgggtgt ggcatatct tctaacaacagcat tcttaacaa agatggaaaa atggacatga 15120	agaacctac	tcttagtggc	tcttatgtta	. cagtataagg	, gagagcagac	tgttccttta	
cacattttat tgacgttagt attttacat gcacagggaa ggagggttt attttattt 15180 ttattttat ctitattta aagagacagg ggtcttgctg tgttgccagg gctggactca 15240 aactectgaa gccaagcgat tettetgett gagatteetg agtagcaggg actataggtg 15300 tgeteetetg tgettggeta aagaaggggt ttgtatgtga ttttaacaa aggetgataa 15360 attgtgaaga agtgactagt caaaggagaa qaggatttee geteecaggg tattttgeea gtgggaagat teetaataag ggacagggaa acacetttae agatggaaat teatateace 15540 tteecacagg gaaattetg teetgeetta ggeagtagg ggaagggeag agaattete 15600 etgtatetge tgggeteteag gtgeetteag eteaaaataa teettataet taacaacage attgggtgt ggeatateet etgatetett teaacageat catetataet taacaacage 15720 aaaggttttt tttaaaaaaat catgttteaa ggtaggaaga atggacatga 15780	cccttgtag	g gtagctaggg	cttgtgaatt	. aagagactga	ı ttaacaggag	aagaggcata	
ttattttat ctitattta aagagacagg ggtcttgctg tgttgccagg gctggactca 15240 aactcctgaa gccaagcgat tcttctgctt gagattcctg agtagcaggg actataggtg 15300 tgctcctctg tgcttggcta aagaaggggt ttgtatgtga ttttaacaa aggctgataa 15360 attgtgaaga agtgactagt caaaggagaa gaggattca gctcccaggg gtggtaaatt gtggggaagat tcttatgtga tatttgcca 15420 gtgggaagat tctctaataag ggacagggaa acacctttac agatggaaat tcatatcacc 15540 tttccacagg gaaattatg tcctgcctta ggcagttagg ggaagggcag agaattctc 15600 ctgtatctgc tgtgtctcag gtgccttcag ctcaaaataa tccttatgcc aaagtagcat 15660 attgggtgt ggcatattct ctgatctctt tcaacagcat catctatact taacaacagc 15720 aaaggtttt tttaaaaaat catgtttcaa ggttgcatg tggaagacaa atggacatga 15780	cacattttai	t tgacgttagt	atttttacat	. gcacagggaa	ı ggagggtttt	atttttattt	
tgctcctctg tgcttggcta aagaagggt ttgtatgtga ttttaacaa aggctgataa 15360 attgtgaaga agtgactagt caaaggagaa gaggatttca gctcccaggg gtggtaaatt 15420 gtgggaagat gactaggaaa tgtatagtaa taaggtttgc tatgcaggtt tattttgcca 15480 gttctggtc tcctaataag ggacagggaa acacctttac agatggaaat tcatatcacc 15540 ttccacagg gaaatttatg tcctgcctta ggcagttagg ggaagggcag agaattctc ctgtatctgc tgtgtccag gtgcttcag ctcaaaataa tccttatgcc aaagtagcat 15660 attgggtgt ggcatattct ctgatctctt tcaacagcat catctatact taacaacagc 15720 aaaggtttt tttaaaaaaat catgtttcaa gatttgcatg tggaagacaa atggacatga 15780	ttattttat	r certatetta	aagagacagg	ggtcttgctg	; tgttgccagg	gctggactca	
attgtgaaga agtgactagt caaaggagaa qaggatttca gctccaaggg gtggtaaatt 15420 gtgggaagat gactaggaaa tgtatagtaa taaggtttgc tatgcaggtt tattttgcca 15480 gtttctggtc tcctaataag ggacagggaa acacctttac agatggaaat tcatatcacc 15540 ttccacagg gaaatttatg tcctgcctta ggcagttagg ggaagggcag agaattctc 15600 ctgtatctgc tgtgtccag gtgcttcag ctcaaaataa tccttatgcc aaagtagcat 15720 aaaggtttt tttaaaaaaat catgtttcaa gatttgcatg tggaagacaa atggacatga 15780	aactcctgaa	a gccaagcgat	tcttctgctt	gagattcctc	, agtagcaggg	g actataggtg	
gtgggaagat gactaggaaa tgtatagtaa taaggtttgc tatgcaggtt tattttgcca 15480 gttctggtc tcctaataag ggacagggaa acacctttac agatggaaat tcatatcacc 15540 ttccacagg gaaatttatg tcctgcctta ggcagttagg ggaagggcag agaattctc 15600 ctgtatctgc tgtgtctcag gtgccttcag ctcaaaataa tccttatgcc aaagtagcat 15660 attgggtgt ggcatattct ctgatctctt tcaacagcat catctatact taacaacagc 15720 aaaggtttt tttaaaaaaat catgtttcaa gatttgcatg tggaagacaa atggacatga 15780	tactectet	a tacttaacta	aagaaggggt	: ttgtatgtga	ı tttttaacaa	a aggetgataa	
gtgggaagat gactaggaaa tgeatagtaa acacetttac agatggaaat teatateace 15540 tttecacagg gaaattatg teetgeetta ggeagttagg ggaagggeag agaattete 15600 etgtatetge tgtgteteag gtgeetteag eteaaaataa teettatgee aaagtageat 15660 atttgggtgt ggeatattet etgatetett teaacageat catetataet taacaacage 15720 aaaggtittt tttaaaaaaat catetteaa gatttgeatg tggaagacaa atggacatga 15780	attgtgaag	a agtgactagt	caaaggagaa	gaggatttca	a gctcccaggg	grggraaart	
tttccacagg gaaatttatg tcctgcctta ggcagttagg ggaagggcag agaattcttc 15600 ctgtatctgc tgtgtctcag gtgccttcag ctcaaaataa tccttatgcc aaagtagcat 15660 atttgggtgt ggcatattct ctgatctctt tcaacagcat catctatact taacaacagc 15720 aaaggtttt tttaaaaaaat catgtttcaa gatttgcatg tggaagacaa atggacatga 15780	gtgggaaga	t gactaggaaa	tgtatagtaa	taaggtttg	catgcaggt	. cattttgcca	
ctgtatctgc tgtgtctcag gtgccttcag ctcaaaataa tccttatgcc aaagtagcat 15660 atttgggtgt ggcatattct ctgatctctt tcaacagcat catctatact taacaacagc 15720 aaagttttt tttaaaaaat catgtttcaa gatttgcatg tggaagacaa atggacatga 15780	gtttctggt	c tcctaataag	ggacagggaa	acacctttac	agatggaaat	. LUALATUACE	
atttgggtgt ggcatattct ctgatctctt tcaacagcat catctatact taacaacagc 15720	tttccacag	g gaaatttatg	tcctgcctta	ggcagttagg	y ggaagggcag	, agaattectic	
agaggerent titaaaaaaat catgtttcaa gatttgcatg tggaagacaa atggacatga 15780	ctgtatctg	c tgtgtctcag	gtgccttcag	toppopool	r catchatac	t taacaacacc	15720
	atttgggtg	t ggcatattct	. ctyatetet	, ccaacayca	tagaagaca	a atggacatga	
ttyayataaa tyaayaatat atatteetta ataaagaatg objections migistory	aaaagtttt	a transparate	. cacyctica:	a acaaacaat	g ctgtatatt	t atgtctctgt	
	ccyayacaa	a Lyaayaatat			J J	_ -	

						15900
gacattgtgt	tatggaggct	aaggtgttaa	gcatgtgatt	actttagatg	ccgtatgact	
acctgttttt	aagattaaaa	aagaatcaat	aggcagttta	tatgcatggg	agcaagttaa	15960
aaacaacaca	gatgtgatga	aggcgaggtg	aaactggtcc	gcatctaatt	caggccttct	16020
cctgaaagcc	agtgtgtgca	agataaataa	gtttgtttga	cgaaagcaga	ataactagtt	16080
tatcctttat	gatgaagata	gttattcaga	aatcatttt	attggctacc	tctgaattaa	16140
taaatgaaaa	gagaaatttt	tttttctgta	ggggatgtct	gatgagttct	taaaaagtgg	16200
atgaacctga	aattatcatg	aacaagcaat	tataatgaac	ttaaaattac	ttaaagagtt	16260
atgaaaaaca	aaaagaaaag	ccgtatgttt	tcttgtgcct	tattttgaag	tgacaaatta	16320
tttgcagggt	acatttgtag	acggaactaa	tgtgatttaa	aaaatgagta	ctagatttac	16380
acaatcatc	ctttaaaaag	tcactggtgc	actttaatta	ttttatttat	gtttattctg	16440
agactacctt	tattttgaaa	atgaggtata	gctttgccta	ctggtgacaa	aagtgtaaat	16500
aattcactee	acatctgtta	aaaaccagct	tagtactaga	ctcttagaat	agaaaactga	16560
tanggantt	gaggagctca	tagtccctaa	addactagga	acttotcatt	aggtgtgcag	16620
teaggecatt	atgctcctga	ageceetaa	acadatacac	accaccatoc	ctggctaatc	16680
tgtgttctgg	tatgtagaga	aggagtgtgg	ctatactacc	caractaat	ttgaacttct	16740
tttttataat	tatglagaga	cagggtetgg	agagagaga	accaccact	ccacctcacc	16800
gggcttaaga	gatcttccct	ccctgcccct	accyactecy	atagastata	caacctttt	16860
ctccccaaag	cactgggatt	gcaggcacgg	gecactatge	tetteet	caaaaccccc	16920
aaatcagtgc	atactcaatg	gtcttgatgc	aattetgget	tgttggtaag	agaacgggga	16980
tttactcaca	agccacgatg	tcacttttaa	ctctgaacag	atcaagctat	tggtattatt	17040
catttatgtc	atcgataaac	tttatgaata	aaaactcatt	gtgcaaatat	ttaaacatac	
tacatacata	gcactgtgca	gtttctaagg	aaagtaatgg	aaacctttgt	cacatccctg	17100
gcttccagaa	ctttatgtta	tctaagtgca	tttgtctgca	aagttgttgg	gttaattgcc	17160
cctttcttc	ttctcttttt	aagatattaa	taaatagtgt	catgaccaaa	agataatcct	17220
tatggacaag	atagatctaa	aaagccttag	ctaatttata	atcttgcata	atccatgatg	17280
acaagatgca	gaaacaaaaa	tgcccagaat	aaaaacttag	caccattagc	agccatttcc	17340
ttttaagtct	ttacaagtat	actcccagtt	tcttgaaaaa	tttattctaa	aatatgtaag	17400
acacacaaaa	cagcagaagg	actaatacag	gtacatcgaa	cacctgtgtg	cctaccgccc	17460
agtttaaaaa	taaactggaa	tgatgtttct	ctcatactta	cagaataaag	ttttaatctt	17520
tagcatggaa	ttcaaaagac	ttctgccatt	ccagttcaga	gccacccttc	tggtctcctt	17580
actecteage	cgcgacactg	cccatqtacc	caacaggcct	ccagggttac	tgcttccatt	17640
cottettatt	ctcatgaaca	ttttccttca	tctcatctgc	cagaatccta	cctaataata	17700
ctcctactct	gcagtttaca	gttctttaaa	attaaaaaag	gttgtgtacc	ctttagtgtc	17760
ctgaaaaaag	aaaaaacaaa	tttaaaacct	taaaaaggta	ccatattttc	atagtatttg	17820
cattatatat	cattacagtt	cctgtggaca	tatctatctc	ttttactaga	ttgattgtgg	17880
actetttaaa	ggaagatata	tettatgaac	agtgtttat	atattgttag	caatcaatga	17940
atacttacta	tatttttctc	atgaggatat	tgattattct	attttaattt	attaccnnnn	18000
nnntatacta	tacataactg	ctttctgtac	ctgagctatt	tatgatetet	gaggeteetg	18060
tananantat	aatttttgtt	aatcatqqat	ggaaatattc	acaacatcat	tcatcaattt	18120
chtanantta	tcttcctttg	tatattacan	atottttaaa	atatcaaagt	aatottttt	18180
cucacatty	tttagatatt	actatataga	gatttgccaa	aaaataaaga	aaatataata	18240
tyttttattt	atcatcaaag	gecacatgga	tatttcattt	gcatgaaacc	taggtttttc	18300
tatttagcaa	accaccaaay	tacagging	attattatat	ttaaatgaag	tagattttt	18360
tacagatggc	acatgggcat	ccadaatacc	getettatat	tattttata	attttacttt	18420
aaaacagcaa	ttttctgtgc	agatattaca	ttttatta	attacattat	scatchatta	18480
ttggaaagtc	agaaacttga	aagctatgaa	CCCCCaaa	taatttatat	ccctctgttg	18540
gatgtaagta	. agctatcttc	ttacttgctt	getttgttt	teettigigi	agctctttaa	18600
agagtgtatt	. cattetttt	gtaagtgatg	tttctagaag	tagcattggt	gggtcgaagt	18660
gtgtatacat	tttacatttt	tgattgctaa	gctgcagaaa	agetgtattg	gtatgtaagt	
actcgtttcc	: ttactatgct	cgtcatttct	agtgtctgct	cttcctttcc	ttcttcaaat	18720
gggtttggtt	taattctagt	tgctactgtt	ccatcagagg	aattgcagag	aactggtctt	18780
caaaacagtg	, cagtatatac	tttaggtgaa	. gatacttcta	aaaacctttg	tattttgagg	18840
taattctaga	gtcccaagaa	tttgcaaaaa	. gagtacattg	tcagcaatat	ttttcccaat	18900
ggtgacatct	taatataact	gtagcacagt	agcagaatca	ggaaattgtc	attgggtaag	18960
gtacttttta	attetecaaa	taattcagco	ctccaaaaaa	. atcccactto	ttatgttttc	19020
aaacctgtac	r ctacttttga	tacatactto	: ctaaattgca	ı tttttattac	: tttaaaaaat	19080
ataataccta	a gaageteaaa	qctqgaaaca	<u>g</u> cctgatcaa	ı tatagtacto	: ttaagctaaa	19140
aacaacctga	tcaatatagt	actcttagg	r aaatcactta	ı tgcctgtggc	: tttttttaaa	19200
ttttcttcct	t atcaactate	tcttcatgat	: tttgtggttt	: ttattactgo	: ttataccata	19260
gatgaggtat	t agaaagtaaa	a agaagttaaa	a atgcattttt	: ctcaatttag	, tgaattaatg	19320
attacattca	a gatttatagg	acaagggttg	g aagctancaa	a ggggttgata	ı ggaatcttga	19380
totatctoa	tattttccc	aactttatta	a catgactggt	t tcagactatt	: ttatctaatt	19440
acatttcac	t cttggcagaa	a atagcaaaa	agtcaacca	a tggtcaatgo	tgctgagaac	19500
	22 3	_				

						10560
tctggcctgt	gcagacatat	tggctgtttt	acttctaata	ccattctgct	tttcctgtcc	19560
tgctgctgat	ggatgtttct	tccaggtttt	aaatatcaaa	caaaagggat	ctgtgggccc	19620
agtacaggga	atggctcttg	atagatttga	ttttcctgca	tttcctttat	tttgatccag	19680
tgttaatttc	atgtagagtt	gtctgtttaa	caggattctc	ttaaaattcc	ttcttcagtt	19740
tacctgccag	cttttctttg	tccaggtttc	agtatgaact	ccactcgatt	aatagagctc	19800
tctagtagtg	acttgtggag	tgggttctct	gaacatttct	ggaagtgttg	ctgatagtga	19860
taatattgat	cactagtact	gttaatttgt	gtgcttacta	catgttggct	tttatatgta	19920
ttccttcaga	ttaaggactt	ctagaaaaca	tccatgaaaa	aacagattaa	aaaaaacaat	19980
totocatota	tttgggacta	gaaggtacta	tgggaaggat	aatcttcata	ctcagaccat	20040
actgacctga	atttcattta	tcagtttaga	gaaccacttc	cccttccctt	caccctacct	20100
ccaaatacct	gtgactttgt	atcaccactc	tagcaccaca	tcctcatccc	agcaggattt	20160
addaadacta	ctttttgaaa	gccttttaaa	attctgtaag	ttgagaaaat	actaggggaa	20220
trattttaaa	tttctttaga	attacagget	ttagtcagta	tatgacagag	ccttttccta	20280
gattttaaa	catataaaaa	tttacatata	attttagggt	ttcagagacc	cctaaagcct	20340
gaaaaatgtg	gtggttcatt	atctaattat	atttaaatac	ccttctaaaa	cccttttgag	20400
atteatagat	tcacaacaga	gtetgattgt	aaatotaatt	accccaaaca	acatttcaaa	20460
atgttaggaa	tcacaacaya	tttetett	atattataca	ageggaaaga	gaataaatga	20520
gactgttgtt	ctgcttagac	ccccagttt	greetergee	aggettgetg	ttcccttatt	20580
gtttcctggc	ctgatactca	aaagaaccga	tatttaaatt	tatattaaat	tatttctctc	20640
ttcgcttgac	acatccttgt	ctctacattc	tgttttttt	staattaasa	catchectt	20700
tcgagtcagc	aggatatagt	ggctgttatt	TCTTCCCCTT	atcottcaac	gacctactcc	20760
tgacaacact	ttgccttttt	ttttttgaga	tggagtttca	ctcttgttgc	ccaggctggg	20700
tgtaatggtg	caatctcagc	tcactgcaac	ctttgcctcc	cgggttcaag	ccattttcct	20820
gcctcagcct	cccgagtagc	tgggattaca	gacatgcacc	accacgcctg	getaattttg	
tattttcagt	agagatgggg	tttcaccatg	ttggtcaggc	tggtcttgaa	ctcctgacct	20940
caggtgatct	gcctgcctcg	gcctcccaaa	gtgcagggat	tacaggcgtg	agccactgtg	21000
ccctgcctgc	tatttgcctt	tttaatctca	tgaaatgttc	tetttettg	gctgaagtgt	21060
cacttttctt	gttgaacagc	atgcgtggtg	agtagaatgt	tataaaaagg	gatggacttt	21120
ggagttagag	agacccaggt	tcctgttcgg	cattgcagaa	atgctgttct	gcaataggct	21180
gtgtgtcagt	gggcaaatta	cttatctctc	agagccttat	tggtaaggtg	tgagtgatag	21240
ctcctttcag	gcaccttaca	gaggctgtct	cctaatcctg	gtagcgtacc	tggctcatag	21300
atggcattta	aaagtggttg	tgatgacagt	catagctcac	cattagcata	gcgctggatc	21360
catggcaggg	aagcgctgca	catgcagtat	ctcttggact	acacagggcc	ctcatgaatt	21420
aggaactgct	gtttcatgag	gatagggatg	aggaaattag	acttgctgcc	cctcactgcc	21480
ttccactcct	ctcctccaag	ttaatgggaa	ctatgactct	gctttggctt	gattgccatg	21540
gaagattctc	acacagccaa	atttattgct	atcttagtta	aattatgcca	gaacacaaaa	21600
tatgaagtta	ttqtcaaagt	aatataatct	cagctgtaac	tgagatagtc	agaaactgtc	21660
tgtaatctga	tqtcctatct	gaaaggtagc	tgagaataaa	caagaaataa	agagaattca	21720
gtagcaaata	ttggtgacac	aaagctttta	tattttgact	agttaagcta	gttcttaaat	21780
gtttccacta	aaatattcaa	gtttaagggc	atagcccagg	gcagcttatt	atgaacatga	21840
tatattttaa	aaatcttaca	ctttctctta	aaagttcttg	ggaggggcat	gtgaggccat	21900
aatataacca	taaaaccatt	tgttttaaaa	taaaacccat	ttttaaaatt	cttccaaata	21960
aaaaaattat	tgcaggaaaa	aatgctaaac	ctggttttta	actttgtacg	ccaactatat	22020
ttccaagatg	tactatagee	tggtaaccat	acagaaccat	acagaattag	ttctcagaat	22080
trattototo	cttacttttg	catttggtac	aggtataaca	gggtcgatta	tatggtttct	22140
aagacatgac	tagaaagaaa	tatgtttatc	agttattatt	tcttccatct	aaattagaag	22200
gaactaggga	gagggcttca	acaggaattt	atatacttta	gagaaaagtg	atcattgata	22260
gcccaatagt	atagatatct	caacccaata	acacaggttg	tgtctgtctc	tgggatcata	22320
cactatagg	gagaatcttt	gcaagcaaca	ttctacttat	agggagccat	aacaaaagtt	22380
tcatatotat	aataattata	agtettaagt	catcaagaaa	aagttaactt	gtgaatgata	22440
atccctgatt	aaaaagagag	atotataata	atggataaga	gatttttctt	ggttaatttt	22500
tagtattaa	a atggctaaat	cttctttaaa	atattctgac	tagtatggtg	cattotctaa	22560
tagtattaa	atagetgaga	actaatcatc	ttgtaatctg	togaaaactg	tectetttag	22620
dtaasactt	atagetgaga attgtaatte	ctctaaatcc	tcagctttta	ttttctacac	acttttttt	22680
tttttsss	tttccttcct	ctgactcact	ccttttatta	tcattttcat	ggcctgagaa	22740
catacatas	gatagaatta	ttcttttcac	agattaacac	ttttctttc	gagtatcgtt	22800
anactanta	t gtgtattaac	tagagaagto	toccttacat	ttcattttta	tattttcttt	22860
gageteatg	a gatagtttgt	accettter	tttcaaatco	aagtttctgc	ggttcttaag	22920
occateagg	a tttgtctcct	. ayucaucac . aaatttaaat	tcatttcctc	tttaaaccat	atcetetatt	22980
accegtate	c tgcacccact	. saaccccact	· ctatttattt	aattoocaac	ggccactctc	23040
tatattatt	a attttttctt	. tttmassmot	. caactaacaa	cttctaggas	gttttttatt	23100
rgrgrrgga.	a tcaattcata	costotteco	cttattttt	r caaccettte	rttaataacat	23160
gctactgtt	a icaaticata	Lactial	. Congressing	,	, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	

atttatttaa ctatagttat tagcagtctg agatcatttt acttggttac ataaggagca 23220 catatateta eccageatea ttgtaaggea tgtgagaeet ttgtttgatt getgteetaa cctagtaccg agtcctaaaa actcattagt agaagatgaa gtgtccttgc cttttgctga 23340 acatatatat acacactgaa tatttagtgg caattcatag ttgcatttgg ccattttttg tttataattt cccctttctc attaaaaaaa ctttgttttc tagactttag gatttagaga 23460 ageteatttt gttecataca catgetgetg ttggattatt taggtatttt gtgaetgtat 23520 tttatctttg aaataaaaag cctttcaaga aatgcaaaaa aaaaaagctc aaaacacaga aaatgtatat tttttaaata totcagatag atttaaagaa attttaaaca tootaatcat agtacttttg aagcccattc atagtacaac ctgtgaagag cctcatgtac gcgctaactg 23700 ggtcctgtct ctgcagttga ctggattgtt gctgacatct tggccatcag gcagaatgcg 23760 ctaggacatg tgcgctacgt gctgaaagaa gggttaaaat ggctgccatt gtatgggtgt 23820 tactttgctc aggtaacttg tttccatgct tttctctcta tatatgtagt ttataaattt 23880 ttttttttt ttttggagac agtctcactt tattgctcag gctgagtgca gtggtgtgaa 23940 cacageteae tgeageettg acctetgggg eteaagtgaa eeteetgeet etgeeteeea 24000 agtagttggg accgtagtgc ccaccatcat gcccggctaa attttctatt ttttgtagag 24060 atgggggtct cgctgtgttg cccaggctgg tcttggactc aagcaatctg cctgtctcag 24120 cctaccaaaa tgctggatta taggtgtgaa ctgccatacc caaccctata aaaatgttat attttaaaat ttaacaatat acttcatgtg aatgtatggt ttttaaaatg ggtttaatag 24240 tttattctca gttgaagtaa ttttgtttgg catttttagt ggtgtgtatt tatatacgtc 24300 tgattatcca tatgcggttt tccttcagca tctgtgggga ttggttttag aaccaccaca 24360 gataccaaaa tctaaggtgt tcaagaccct catatagaat gggatagtat ttgcatataa 24420 cetgtgcact actttaaatc atctctagat tacttataat atctaataca ttataaatgc 24480 catgtaaatg gttgttatac tttattttt atttgtatta ttttaattgt tatattattt 24540 ttaattttta tttgttcaca tatttttgat ctgtgatttg ttgaatctgc agatgtggaa 24600 ctcatggatg tgaagggcca gctgcagtaa aatgaaaagag caaaaatgca aatgtacaaa 24660 gttcaaacaa ataggaaatt taaaggcata gaatttgata ggcaattaca ttaaactgtt 24720 gataacagta attagtgatc tgtatgatat taaaaaaaaa aagcaaactg tatatataaa 24780 acttactttc tccagttctg gaggctagac atccaagatc aaggtgttga cagggttagt 24840 ttctcccaag gcctctctcc caggcttgca gacagcatcc ttcttcctgt gtcctcaggt 24900 ggtttttttc cctgtgccca agcaccctg gcactgcttc ctcttcttag aaggactagt 24960 tacactggat gactaatcct totacagaga ctgctaaggt cocactctga ggcccttttt 25020 taaccttaat taccacctct aagtccctct ctctgaatac agtcacagtg ggaactatta 25080 gggctttagt agactgattt gggggaacac acttctgtcc gtaacagtgc cacataaata 25140 tetttagcag gattgatttt ttaaaatccc taaagatcgt gagtattgac atgttaagga 25200 cgctttttag tgactctgta ataagtgggt ggaagaattg ggagttaaat ccatctgatg 25260 gatcaggttt tttatttta aaaatgtgta tttaagaaag aaagcatttt cattttaact 25320 gccaacaaaa ctaaacttca tgtgttttcc aatacagtgt cacatgcagt ttttttgaat 25380 tatgttgaga caaggcaatt ttcagctaaa tgttctttag aagctaatgt ttgaagatat 25440 taaatataga ttaaattctg aaatgtagtt ttcattctgt actttttgca agagaagttg 25500 cetttttgat gactetggcc aattgttatt ttaaaagtaa atgetettte teeegatttg 25560 attgtggcag catggaggaa tctatgtaaa gcgcagtgcc aaatttaacg agaaagagat 25620 gcgaaacaag ttgcagagct acgtggacgc aggaactcca gtaagagcct acccgttttt 25680 atttttctta ccagctctca gtttctaaat ttaagaatta aattaaaatc taagaattgt 25740 tttgacaatg tattttccca tgtgtaatta ctaattcagg gttatgctga ggtaacagaa 25800 accetetatg tacaggtagg caggttttte agceateaga aagattgetg taaacaacta 25860 ggtcctttgc tggtcagtgg accttaaaga ggaataaaaa gagcatttgg tgtcgttcag 25920 agtctataaa tagaactaac tgcattttaa cctgacattt aagctagttt acaagctcat 25980 cttacttctt gtcttcttta gtatcagatt tggttttaga agcagcaact gttttctgtt 26040 agtgcaaatt ttgaatgtct tacatgtaca gaaaaaccaa aaaaggatga atctctacaa 26100 atgttaaatc attcagtgta aataatatt tataaaactt tattccacaa aagtggggag 26160 agttcaatct gctttgtata gaatgctgat tgctgccaaa ggcttttccc ctggttccct ceggagacaa agcaccatga teacegggge gaettggget ttetetttea gtacatgaca 26280 tgtgctcaga agcttagctc gtgtgcacag gctttccctt tcctttctgg ctccctccct 26340 ctgtcttccc tcctctcctc ttgccctccc ctcaccaggg gtcctgggca gcagctggag 26400 ctcatggtga aggaagaatt cttcatggtc agctggcgaa gtgcctggtg tgagcattgt 26460 ttattcacat gcctcttcta ggtgttttta cattagaaca ttgcatctgt tttgggcatg 26520 tgttgggtga cagaagcaga atggaatgag atgaacagtg accetttate ctgttatage 26580 taaccettga gaaccaaget tggtgtette aaagggtetg tttagtetga aacagtgtgg tgaatttggg cagaattgtg gtcattgcat gtaggtctcc aaaagacaga ataagttggt 26700 aatatggttt atcgactttt tacaaaaaaa atttaaaaat catgaattta taccttaaaa 26760 tgtccatccc acttctctcc cagctgtcca gtcaccccag caatggatga ctgctgtgga 26820 gtteettetg tgteetgetg tgggeattgt atatatgaag caaatgaaga tagetgeett ttgggtgatg ttggcatcct atgcacagtg gtcccttgct tttttgcccc catgaatata gctgccagtg gcgctagggc tgaaaaaatc agctctttac acttgtcatg tgtcttgttt 27000 atgtggctgc cttcgtgagt ttcttcttgt ttttggtttg cagcagttta agtatcatat 27060 atctgagtgt catttaaaaa tttttacctg gattggtcct ctgagcttgg atctatgatt tggtgtctgt tattaatttt ggaaatttct ttgctcttat ttccttaaat attattccta 27180 ccccagtctt tettetecag ttatgtttgt gttggtteat ttetegetgt tetttagtte 27240 ttagatgcat tattcgtttt ttgttggttt ttttttaaat tttttttt acgccccctc 27300 cetttttet ttttgtgtta cattttggat aatttetgtt gacccacett tgagttcatg 27360 gattetteet ttggetgtgt tgagtetact ggtgageeag tttaaggeae tetteatete 27420 tgctactgcg tgtttcattc ctcacatttc cctttgaccc tgtttcatag tttccatctc 27480 tgtgctagtg tatctatctg atcataaagc ttagtcacgt tttccagttg aacctttatc 27540 attttattat acttgcagtt ctcttaaatt ccctgcttga taattccaac atctgggcca 27600 tatctgagtc tgcaaatttt gattacttta tctcttcaga ttgtgcttta tcttgccttt gtcatacttc ctaagatttt gcctaacgct gggccttttt tgtaagacag gagaaatgga 27720 ggcaagttgt cttgatacct ggaaatggat agacttgtct ttctgcttgg cttttagtgt 27780 tgaggagtgg agtcagtcca ctgaggaggt gcactgcatt tgggttttgc tcatgtgctt 27840 tttctcacag cttcaggttt ctgtagaact cattactttg tttgtaggtt ggggatgtcc 27900 tecegetaga getttteete agtgtetatt teacaeteag egtttteaca tageaeettg 27960 gagtggetet ettetttatg cettteccca ctatacttet tggatacttg ttactgaact 28020 ctcgctagtt tggtggtaga aggagggga agggaagtgt cttttcattc ttagggagaa 28080 teteaggggt ggageettet etgateetge ettgettetg getgtaagte tgtgeecagt 28140 atgtatteet geetttacta agagttttte cetgttetet teacceagee teategagta- 28200 ttcatccgtg ccccatgggt agcagggttt tgttgcccct gttcatcagt ttcaggctgc 28260-ctgcttttcc ctcccacaag cctacatcca gtcttccctg accgcagtgt gttttctttt 28380 ttetttgtet tgfgagtaca caggaggtet gtgggtegag cetgtgaaat gtgetgeatt 28440 ctecttgtgt etgtagecca ggggttegte tgttecaetg geteataett ggetttetge 28500 aaaattgata aaatttttag ctaaattctt tttactggta tctgttacat tggcccccaa 28560 ctaaacaacc acttgcatct tgtttctcct ttgagttttc catctttcct tagacttttg 28620 ggttagttgg ttgccttgca accttgcagc tctctgaagg gtctaagaaa agtcatgaat 28680 ctacagettg teagtgttgt tgttgttgta gggttggeag tagtatteet teageattet 28740 acatacttaa tggaagccgc ctcccatttt tggttaataa atttcaaaac ttggaacaat 28800 gttagattta caaaaacgtc agaaagaaca gagtgttcct gtttattctt tatatagctt 28860 ttttttttt ttttttttt gagttggagt ctcggtctgt cacccaggct ggagtgcagt 28920 ggcacgatet tggctcactg caacctctgc ctcacgggtt caagcaatct cctgcctcag 28980 cctcctgagt agctgggatt acaggcgtgc accgccatgc ccggctaatt tttgtatttt 29040 tagtagagac agggtttcac catgttggcc aggctggtct cgaactcctg acctcttgat 29100 ccgcccgcct cggccccca cagtgctggg attataggtg tgagccacca cgcccagcct 29160 tetteateta getttaacat etaatgttga catettacat aacatggtat atatttgtca 29220 aaactaagaa ataaacattg gtaccacact attaattgta ctacagattt ttattcagac 29280 tttaccaggt tttccactaa tgtccttttt ctgttctaaa atacaatcca gaatagatac 29340 aaatccattc aacttcagtg tittaaatta ttgtttttca ttatatgaag tgctgtgtgg 29400 tttttgtcaa atctgttatt ttggttttaa tcttcaagct tgtctttgtt tctttaagtg 29460 ataaaggcat aatttaaaag gtgtgttggg ttatttcagt gcctaaagtc ttgtctgagt 29520 cacttgtttt ctgctgttct tgcttatggt actttctttc cttgtttgct ttgttatctt 29580 cetttgetge tggetgtgtt tggttaagtt atttgtggaa atcagttgaa geetcaggtg 29640 ggagtgtctt tctccggaga acatttctac ctgttttagc tgggcccctt aaggctcctc 29700 tagcgtgggc cccacccaaa cgagattctg agttgaaggt gaactgagcc attcaggcag 29760 tgcagccagg gttgcagatg cacgtgagac ctgctcacct ctcatttact ttcaccctga 29820 gagtagagec triggtgttt egtteacttg tetgattete tetteacagt tetattagaa 29880 ggtccatggg ttttggtttc tgtgcccttc atcttatgag tcttgtaaat caaagttctg 29940 ttttatgctt acttctgctt tactgtgttt gcttaatttc agtcttaaca tcttgccaac 30000 tottgggtac ttttaaaata atgttatatc cagcttttta agttgttttc agtaggaagg 30060 ttgattcaaa taacctagtc tggttatggg ctacgagaat agcctccctg ttttttgtgg 30120 gcaaaattcc agccttttat gttcctagcg cagtgtggat aacagactgg caggttcaag 30180 aggeogtget gageagettt cactgtaagg teactgteee aggtegggtt tetaagaate 30240 tggatggttg tttcatttct taatatgtac gccctgtgag agcggataca tcttgctcag 30300 gttcttatga ttcttttgtt tctgaaggtg aattaagtaa gtgacatggt agaatatgtt aagtcaactt tcgtgtggct tactagttct catgaatcta ttccatgatt gtatcagttc 30420 ttattcagta ttagtattta agaaatgcag aattttgttt caaaaaatat atttgtatta 30480

						20540
taagttgtga	agaaatacat	ctccataatt	attgctggga	caatacagta	ttttcttaag	30540
gaacttattg	gttgtggatg	caaatgaagc	atatttgtga	taaaaataac	taatagaagt	30600
cattttgtta	gactatgagc	tagtaaaact	tatggcacaa	acatggagac	ttaacacttt	30660
ttcttccagc	tttcacttaa	gttccttttc	agataggagg	cagcctggtg	gataagagta	30720
ttggttttga	aattagattc	aggtttaaat	cccagatctt	ctgtttaatc	tttattttat	30780
ttcaggtaga	ttttctggat	aacttgctat	agcttatacg	tcagtacttg	ccacttcaat	30840
tttatgttat	ggagagacgg	cttctttcct	taaacctcac	gaaccaacct	ctgctagctt	30900
ctaagttttt	tcctgccact	tctttacctc	tctcagcctt	cagagaatta	aagggagtta	30960
agaccttact	ctggattagg	atttgcttta	agggagtgtt	gtggctggtt	tgatgtttta	31020
tctagagcac	tcaaactttc	tccatatcag	caataaggct	gttttgcttt	ctaatcattc	31080
argrattcag	tgaagtagca	cttttaattc	tctttaagaa	cttttccttt	gcatccgcaa	31140
cttaactatt	tagtggaaag	gacctagett	ttgacctacc	ttggctttca	acataccttc	31200
ctcactaacc	catttctagc	tattgatgta	aagtgagaga	catgcaactc	ttcctttcac	31260
tagaacactt	agcagccatt	gragggttat	taattggcct	aatttcaata	ttattatatc	31320
tcacccata	gggaaaccca	aggaggagta	gagagaaaga	gagacaggag	aacaggccat	31380
cattgggaaca	gtcagaacac	acacqacatt	tatcaattaa	atttgtcatc	ttatatgggt	31440
caccygagea	gcacccccaa	acaattacaa	tagtaacatc	agagatcaca	gatcacaata	31500
geaacteatg	taatatgaaa	tatteteada	ttaccgaaat	atgacacaga	gacgtgaggt	31560
acayacataa	tgttggaaaa	atrocaccaa	tagacttgct	caatacaaga	ttotcataaa	31620
gageacatae	gaaaaaaatg	caatttccct	gaageteagt	aaaacaaaac	atgataaaat	31680
cetteaatgg	tgtcactcct	aagaatgttc	ctotacaaot	tttstgcatc	tottacttac	31740
gagatgagee	tgtcactcct	tatatttt	gagtacatgt	atttttat	ttttatacat	31800
cttttcctat	ttgtgaatag	accelette	gagcacgcgc	tattgatttt	tttctctca	31860
ttatatgtat	cttttgaaga	atatttaaa	adycctaacc	ancttaacc	ttactaagat	31920
taatttccac	tttttgtatc	Ctatttaaga	ageceeegee	thtogathta	attatagat	31980
tttctccttt	gttttcttct	ggaaatttta	gagttttget	ttacattta	tttaatataa	32040
ttatttataa	ttaatgtttt	catatggtgt	aagategaag	tesastaset	agactasets	32100
gtaaccatca	ctatagaaaa	gattatttcc	ccccaatgtt	tgaaataagt	agactgaata	32160
tagatgggtc	tgttatccct	agatcaatgg	agcactigit	ttaatttaa	gacccacaca	32220
tatatatcct	tatgccaata	ccatactgtc	ttaataatge	cigottigea	graagette	32280
aaatagtgta	gttgtcttct	aaatttgttc	tttettttea	aagttgttt	ggctactta	32340
ggttttttgc	atttctgtgt	gaattataga	attagetega	caatttctac	ttestatase	32400
tgggcttttc	attttgattg	tattgaagat	aragargaar	ttgggaagaa	tigatataat	32460
aggattgaat	ctttggattc	atgaacgtag	cctgcatttg	tttacttagg	cottotttat	32520
ttatctcagt	gtgttttgta	gtttaatgta	cagatttgca	catcttttgc	cagatatate	32520
cctaagaatt	tcagtttttg	atactattgt	agatgacatt	taaaaaaatt	ccaagiiiii	32640
gtttgttgac	ctaggcatat	atttgacttt	ttaatatact	aaccttgcta	aacttattta	32700
tcatctagta	acttacaaaa	tatattcctt	aggatttcct	acataaacaa	teatgleatt	32760
gttttagaaa	taacagtttt	actttgtcct	ttttaatctt	gatggctttt	atttetttt	32820
cttgctaaat	tttctggcta	gacctcctag	tacageettg	actagaactg	gtgtgaggga	
aatcctttcc	atattcctca	tctttaggga	aaagcactca	ttctttatc	cattettag	32880 32940
ttcctagccc	cattgccctt	cctaaatttt	ttctcatcat	tttccttcat	cacaccttgt	
tctttttctt	tgcaatcata	tcatgatatg	taacgacatg	tttttattta	tetgtttaat	33000
gtatttcttt	tecteacttg	tccatgaagg	gaaggaccat	atgtgttgtt	accettigig	33060
cagttcctgg	aacataataa	gtatataaga	aatagtttct	gaattagctg	tgaatgaatt	33120
catgccttcc	tgctgtctgt	caatgttctt	ttaaattaaa	catctaagac	agcaaataat	33180
accacatgag	ttattaacct	gagaaataat	cgttttattt	ataaatgact	gagttgaaag	33240
ctgatagccc	: acagtaattg	ctttcatggc	tttgaatata	aaccttactg	ttacaaaaca	33300
cattttcatg	, aaaatgaatg	tgtggtgttt	ggaactagct	ttaatgtttg	tcttcctgtt	33360
tttccttcta	ı gttgctataa	tataataagg	aattttgtat	gtttttccta	attgtaccca	33420
cttttctaca	a ttttcttaac	agatctggtg	aatcttcatt	attaaatata	attatacata	33480
taaattatto	, tttaataata	atattaatta	ttaaaaataa	tataaattat	taaatataaa	33540
gatacatata	a atattatctg	ttaatttcta	. agttaggtgt	gggttctgaa	gactattata	33600
tgaatgaaca	a aaaagcttgc	atatttgcgt	ggaagctgaa	. agtacgaaat	tttagatac	33660
cattatacca	a gtatctaaag	aaaaaattca	gtaccacata	. ggtttttaag	taggagctgt	33720
atgatcata	gtcatccaga	tgaaggaagg	cttctgtacc	agacgtacag	aggtagacag	33780
tattatcta	a gtactgtctg	agatctggca	agaatgaatc	: caataaacgt	agttttctcc	33840
catgagete	c tatcttattt	cctgtattct	gtttgtattt	. gaaaagattt	ggtgtgcata:	33900
acttatttt	t gtcttttggc	: tgtcaatcaa	agttattagt	gtagtttttg	taactcagtt	33960
ctcaagcta	g gagtttttg	tgtataattt	: taatgtttct	gtttttactt	: tcctaagcag	34020
ataagcgta	a aaacttagad	: taattgatta	cttattaaac	gtccagcttg	atattcttct	34080
ttatattat	t ttagtttcag	g tttatataac	aaatgaggtt	tcttataaat	: aaaatttaaa	34140

					+ - + + +	34200
atgcactaaa	ggagctgtgt	gaaataggaa	ttctgtgtga	agettttgaa	cycyaacacc	34260
tagaacgttt	cacatggtgg	gaatttacta	tatgattttc	atcaaatgag	gtacttttta	
gtgttggtac	ttaacgatac	tgatttctaa	aatttgtatt	tctaaaaatg	acgtattaca	34320
ggatctgaaa	gggcaaaaac	tcattgaggc	tttgtatgag	tcagcgtttc	atggcctatt	34380
tttaattagt	gaattattag	catataatta	gaaatgtttt	tagattcttc	atggctgacc	34440
taccaatgaa	totagcacto	catttaaaat	atagttcacg	ttatgttcat	acttaattgt	34500
tgcattttgt	ttgcccctct	tgaaacgaag	gtcacatgta	aataaatata	cattttctcc	34560
tactgtagga	aatactctgt	tagcattagt	aggtttagct	tttttaggtt	aacaataaca	34620
222222334	tcacacaaaa	taaaccaaat	ttgctctatg	tcccacagat	gtatcttgtg	34680
atttttccac	aaggtacaag	gtataatcca	gagcaaacaa	aagtcctttc	agctagtcag	34740
accettectag	cccaacgtgg	taagtaaaaa	tttgagtgtt	tgaacaaata	attttcaaag	34800
geattegetg	ttttagtttt	tattactaa	aaacatactt	ttattttaca	gttgaaggaa	34860
ataataacat	ttttagtttt	attactor	atattatete	ttaggataaatg	aagtttcttc	34920
tgaatgtatt	cattccttga	actagigiac	totatete	anatranatt	gaagtgettg	34980
tccttaattc	actttcatgc	tattattaca	tatatetgag	aaactaagcc	catrattato	35040
ttacgataca	tattcttgtg	ccatggattt	atttaaaatt	tatetaagta	catgattatg	35100
tagatggaag	ctttttctac	agtgtatggg	ttatatgtaa	tggagettet	gttttgtaag	
atgacagacc	taagttggag	tccaaactcg	tacttttatt	agctgtatgg	ttgcaacttg	35160
gaagttgtgt	aatgttgctg	agcttgcttc	ttcatctctt	aaaagaacat	atgccttata	35220
agtagateta	aatctgtgtg	aggattagat	tagaaaatat	gtcaagtttc	tattggagaa	35280
gttacacaaa	gttggtccac	agtgcttgga	agctgttaat	gtcttcaaca	atggtaatgt	35340
tcttaatatc	catattttag	aaaattgaat	aattggtaca	ccaataagct	atgcaattta	35400
accaaattgg	gaagtataca	gaaaacagtg	gctatgctat	gttcttagag	gtgtctttga	35460
accttgactg	tgatttagtg	tgtgatctcc	atatottoat	agtcactcac	tgagcaaata	35520
agettgaetg	gacattacag	cadddcctat	gacagtgctg	totaatogaa	ctttctgcaa	35580
technique	gttcttcatc	tattatata	antatactaa	ctcctaccaa	tataatttt	35640
taatggtaaa	catgtgacta	atactact	ctaatttta	attttattta	attttagttt	35700
gagcattcaa	catgtgacta	grycargaaa	taattacaaa	accedateta	tcataggtat	35760
aattaaaaat	aagggggagt	ttttacaagg	cgcccacaag	tattttaaa	cacatatact	35820
atgacatcat	ttgtaacagt	acttttaaaa	aatyccaytt	teteeseste	tacatgetet	35880
attaagtaag	gagtgtttca	gaataggagg	geccageegg	attatatat	ttaaattaat	35940
cttttgactt	tcattgcttc	ctctgtctaa	tagacatgac	gttetgteat	ccagcigci	36000
cttttgcaat	gccattgtct	cttttgccct	tttcacattt	attaaacaga	acaaaacaaa	
aaccactctc	gaatctgtag	tctacctttg	ttgtaagcac	tttttccagt	acteactetg	36060
ccctcaattt	gttttggtct	gatttgaaat	tctctcccta	gacttctgtg	gggctgttct	36120
ccattatcct	cccaactctc	tggcgattac	ttcctagcct	cctttccagc	ctctttctgc	36180
ttcatttctc	: cctgctacat	gtgttatttc	cagtgtcagg	ttttggtgtt	tgattaattt	36240
cactttttqt	ttctcatggt	ggccttcctc	taaatccatg	gctttagcca	tegttteett	36300
gactgctgat	gactcgcaaa	agcttcctcc	cctccatgtc	tctctgccta	actctggacc	36360
catttgtaca	attgtccatt	agagagette	gcttgactgg	cccaaaagga	tgtctcaaac	36420
tragratati	gaagatagaa	tttatccttc	catgcataca	ctcatatttc	ttgtcttggt	36480
aactccatca	a ttcagttttt	ttgcctaagt	tttattcaca	aaaagaacaa	attgatagca	36540
attacatac	tcttatagga	aacttagaca	tggaggaaga	agctgttcag	atggggtcct	36600
geegeacac	aggcactgtg	graatattta	aacttttctc	agctgttcga	agggttttgt	36660
gcagaagcg	t tttccttaga	cttattttaa	gtatttggct	ttctaatggt	tataagggat	36720
tttaactaa	a atgtatctta	atetecases	tagacccatt	aaadtaadcc	cctatggtgg	36780
gtggaatta	a atgtatetta	accigccacc	atacttcate	acassaatas	catacttaac	36840
tttttttt	t taattgccat	ggttaaaacc	atayttytta	gegaaggega	taacaaaaca	36900
ctttttgaa	c tctcttaaaa	gaaaacagaa	atttaatgat	gigiciataa	tagaaatta	36960
gatacctag	a atttccatgt	tattcatagg	gtgaataaca	ctggcgattg	tagagattty	37020
agagttctt	t caaaacagga	gaacaaaggg	aataagctac	aaagcaattt	ttttttttt	
agacttaac	t gaataaaaat	tatttttatg	tctcaaacat	catatgaaca	aatttagttg	37080
gcaaatggc	a agctaataat	attttataat	ataggatatt	. aatatactta	atattacaaa	37140
agtgcttca	t aattagaaaa	gacataaact	agaaaaatgg	gaaaagggca	tgaataagaa	37200
attcaagag	a tacaaatgac	ccacacactt	. gaacaaatgt	: ttattctttc	tcataatcaa	37260
agaagtaga	a attaaatgaa	tactttqaaq	ccaacttctg	_r agaaagcata	gcaaacaaga	37320
aagctagtg	c tcagctttgt	gtggtaacgc	cactctcgct	: cttaagaagg	tgtgtttgct	37380
ccctataac	t gctctcaggo	agggccacaa	acttaataac	ttaaaacacc	acagatttct	37440
tetettaca	t ttgagaagto	tgaaatgggt	cttactcago	tgaaatcaac	gtgttggcag	37500
gactacaat	c ctttgtggag	acttagagag	atcttottct	cctgtacgg	gtcctgtgct	37560
taattaaaa	g teetgtgett	. actotagagas	cctatactte	gttcgaggt	ctgtgctggg	37620
tagetceggg	c tgcttttaco	accttcaact	tcatctgga	atggcactg	ctccccaca	37680
cccagegee	t gactctggtt	- ctccctcct	ctcactcact	: ctaaacctd	attttaact	37740
ccatatago	t gactetygtt	,	tacaactta	agggeettet	gcagctcttg	37800

tttcctcagc ccttggcaac cctgtttgtt ttctcccttc cttagcagtg gagaacatcg 37980 agtotgtgtg gagcagttot tttatttttc tccttttgac tacctcatgg ttttcacgga tttttgttct cttcacattc aaggattttt tgctttcaga aagttatatt tctctggaaa 38100 gagtgcaccc aatatccctt ttgatttcaa aatcttaatg tggagtctct tgacttggat 38160 ttctttggaa gaaactgctg aagctgccat gtctaagaag aaaactttgg agaaaaattt 38220 tottottaga catggcaacg tcaacagttt ctaagctctt gattccgtct accetgtctc categitiges teagiteatet gestlactic tetigeagggg titteteceag citigeaaatg tactccaatt ctgaaataac taagtctata gctgtgcaaa gagaagtctg ggccccttgc tttcttgtgt ttgactccat ccactctcca gaaatgaatc ccacttctca cttaaccact gacetecaaa geategtate atttgtgtea gttgteatat ttgttaaett teacataaet 38520 tttgacatta tttatacctt tataaccagg aaataatttt aactttattg tagaaataaa 38580 caatggagta taatttttct tgttgaagat aaatatcacc tcctcttcct ttaaacatct 38640 cttccctttg tttttgtatt acattggttt cccccctttt tttatttcct gggttgtcgt attccctgtt attatttta ccttttttt tttaatgtgg atgtttccgg agtctgtatt tottgccttt tcatcttctg ccctttatta ttctcagcca ctgccattac ttcagttatc cattcccatg gtttccacat gcttagcttc ggttgattct tgccatttta cagaccatat 38880 ttccaactac ttctagaatg ttttgttcct tcagcctcag tatgcccaat ttgaactcat 38940 gttetetete eccettett etteettett tetttegete teteteeett eettettte 39000 39060 tttccctccc tccctttctt ccttccctca ctcgttctct cttgcttgct tgctttctct cetetete ttttettet gennnnnn nnnattette tecetecete tetteettet 39120 ctccccact ccccaacttc caggctaaag cagtcctcct gagtagttag gactacagac 39180 atacacgtgc caccgcgccc agctccgtgt tctctttgtt tccctgcctc ctgctcttcc 39240 acttatettt geatggeagg tgggtgeaeg eaggeatget etgeatgtet teetettgge catteceett ctagttatgg tgtggettta tetacgegtt etggageaga ageetagtea 39360 caaagctatt tttttaaaac attcatgata attcatttcc ttttatgttt taaaaatact agetttetgt etttatttee ttactaaett aettggatge eagtaattag ttgttttagt 39480 gaacaccaca gagtgatatt ttgaaacttt ggacttcata aagttggatg agctccagta gcaaagaagg aagtgttaac tagtttaact gacaaataaa tgcttcccag cttggtgtgc 39600 gattgagatt tttgttgcaa gtttgtgaat caatttaact gcccctgccc tggggactaa 39660 agtcagatac gtgcttgtgg gaatctttgt ctttcccaca ccaccctgca ttttaaaacc 39720 tcttgtgtgg gacagtccca ccatgtaata gctgttcttc cttactcagc tactttccct ccagagaggc cagtagaaaa tctagactag ttttttatag tctattttca tgtcacttat 39840 39900 tgagagctac tgttttctgt taaattgtca gtaaatattt taatcaagga aaagggaggc 39960 aataggaagg agagaagaac aaatccttaa ccctagtagg aacctaatga atgggatttg ttctggataa ttgcagtagt cccccagcta aagaaccttt taaaaatatg tcagatatac 40020 ccaagaggat tgaaatcgta tgttcataca aaagcttgtt cacctgcagc cttcatatgc 40080 aattoctatg aatgttoata goagoattat toataatago caaagtatgg atgcaaccca 40140 aatgtccatg aagcaattaa taggtaaaca aaatgtgatc tgttcacaca gtggaatact aactattcag ccataaaaag gaatgaagca ctgagtcctg cagccacaca gatgaacctc 40260 agatccatgc tgagcgaaag aagccagaaa caggaggcca tgtgctgtgt gactgtattt 40320 40380 ctaggaaatc ttgagtcacc atgggcaaga tgctatcacc tttgttcagt ggccagaagc gagggcacta atatttaccc ttgccggggt ctactagatt gaagcgtttc cgctaggcca 40440 taaacttcca acacggtgac ttgtacatgt agatatttga tcaatatata gcaaatgaat 40500 attgatttaa acagaaaaag gcaagtgaga gtgctttcta aacttagagc cctaaatata 40560 tgaggttgtg gaattaatag attctgttgt gtgtgtttga gggaatttaa aaataattta 40620 gatgttaaac agtatattgt ggaggtgttt tgtaactaat taatgacggc actgaattga 40680 cttctaggcc ttgcagtatt aaaacatgtg ctaacaccac gaataaaggc aactcacgtt 40740 gcttttgatt gcatgaagaa ttatttagat gcaatttatg atgttacggt ggtttatgaa 40800 40860 gggaaagacg atggagggca gcgaagagag tcaccgacca tgacgggtaa gtgtgttcac gcacctgaaa tgcctgtaca cggtatatac agtgcacatg tttatgtaga attcagtttt 40920 acaaagtagg ttaagtgtac tttttcctc cattacattt acccggtata tttttcaaga 40980 tgttattaag atgtaacagt ggagatttca ttagtcctgc aaagtgtggt atttcttggc 41040 tgtcgtgtga gtcctgtgga ctcaccaatt atcattaatc cagcctcttt ctactcaaag 41100 ttcacactta aaaggaaagc tctgtaaaag ggaggaagac gtgaagaagg agcacgcctg 41160 gcagtactga gtgcacgtta ttagtcagtg ctgccctttt gctgtatttt tcgtaaaata 41220 tttattaaat ttgggtgtca ttgtgacaag aagaaatgca gttaagtgtg acctttttt 41280 ttccccaaac atgttaggtt ttaagaacct ttgagctatt gtcagatata accagaaaaa 41340 aatagaattt taagtgagca ggataactta gttaaactaa ccaaacatag tgttagctgt 41400 tagagaaatg taaacatgga aataggcaaa cagggaagtg tgtggagttt ctgtttcctt 41460

	ctgtttgagc	taaaattaaa	agagagact	aggetteatg	gagtttttt	41520
ttcaaaatat	ttttgttttg	cggggccgag	ttcactctat	cacccaaact	ggagtgcagt	41580
gtttttcgtt	ttttgttttg	agacaagagc	ctcgccccgc	cacacdattc	tectacetta	41640
ggcgcaatct	tggctcactg	caacccccgc	tessesses	catacgatec	atttgtattt	41700
gcctcctgag	tagctggaac	tacatgcgtg	coccaccatg	tangactaat	tacctcacct	41760
ttagtagata	tgggatttca	cettgttgge	caggerggre	tagataataa	actatttaaa	41820
gatccacgca	cctcggcctc	ccaaatgagc	tetgtgtttt	atatactasa	gacacatatt	41880
gttgagccac	tatgtatgtc	agtgtgcttg	tatcagtagg	acctactgag	ttaatataaa	41940
caaaatatga	gcctccagca	cgttttacat	ggaaaccctc	accigaagca	cogecegaa	42000
gttgatgtgc	cttggaaatt	ttatagagta	atatttttaa	Ctacaacaaa	acacccaca	42060
aagtagacat	tattaaagca	ttcagaagtg	agcaaggata	gaaattatte	Lgcccaacct	42120
tacacgtagg	ccttctagac	gtagtactgt	gcaccgttac	attatetaae	tatassatas	42120
tgtcatcttt	ggatgttagg	gatttttcca	aagttcagtg	agattatagt	cyccaaacya	42240
ttagtctgtt	aaataatgat	aagatgaggg	tcactcaggt	tttaaaagaa	aagetetteg	42300
actgaaagag	agagcagctg	tctactgcag	aaagttaggg	agggaggccg	gaggagtgag	42360
gcccaggggc	tagctagtat	aaaaattggt	tatggtcgaa	ggaaaaaaaa	atgtaacata	
tttatatctg	aaagatgatt	gttctcataa	ttgtatataa	cacagagtaa	ttgtaaagta	42420
gaaaactaag	gtgtttttca	ttttagatgt	aaatgtttag	aatatgtaat	gcatcagttt	42480
aaaaattaaa	actotacoaa	atgcacagtg	aaacgtcttc	cttgctttcc	accctgctac	42540
ctaaccttcc	cttctccttc	ctagcgataa	ccagttttct	taatttgttg	tgcgttgtat	42600
gtgcaaattt	aagtatatct	tcttattcta	ccatccctcc	cttcttacag	aaaagtggca	42660
rattaatatt	tttctctttt	aaactatcga	aggagttact	tacctatttt	tgcatttcaa	42720
aacagacagt	tcatcaagat	tatcattagt	ttattaaaca	tagittaaga	ttaaacaagt	42780
gtttataacc	aatgaaaaac	agatagactc	cccataataa	ccttgtttaa	atgotgotac	42840
ttttatcatg	teceetecta	tctaagaacc	ccttggttca	gcagagctca	tgggtaaggc	42900
cageetetgt	tacctaccat	cggaggaatg	cgttccagcc	gtgatctctg	ccttgccttc	42960
acttcctcct	atactatacc	gtgaagcctc	ggccgtggtg	aagctggctg	actgagtcct	43020
cctgcacccc	atocatatto	agtagttgaa	ggctttgtgt	ggccaatcct	gctttccaca	43080
ggaaaccacc	ctctcttttq	ttgccctcat	ccaaggctac	tgttctccca	gagtgacagg	43140
cggcaccttt	cccagcatag	cactgtgcct	tctcctgccc	ctgctcttgc	agtactgctg	43200
taacactaat	nacatatatt	acagtgctgg	cacttagcac	agggctctgc	CEEECECE	43260
teccageege	atcataaqtq	ccttgaggaa	gccaaaacct	tctytgagtt	gcattgcctg	43320
ggttccaacc	teccaetgee	ctgcttatcc	tctgctacat	gtgagctgac	tgtggetttg	43380
gaataatcac	tacctatata	tattcattac	aaattgtctc	cttttgaaag	attgaccttt	43440
ctgacttacc	cagataccat	aaagaaaata	aaatcttatc	acttcagtca	aggataaagt	43500
atttctgaat	taaaggaaaa	atacaccaga	gtaaaatcaa	gactgaaaga	caaactggga	43560
aattatttgc	aacctagatc	atagaaaagg	ggtcatttcc	ttcttgcgta	aagtgcactt	43620
acaaattgat	aagaagatga	ctgataacta	gaaagaaaaa	tgggtaaaga	acaacaatay	43680
acatttcaca	tttaacctca	ttcatgataa	ggtaagtgca	aatgaaaact	acaggggata	43740
ccttttttt	tttttaatcc	attagattgg	caaacatccc	aaggtttgat	cataggctca	43800
gtgggtgaga	tttaagtatt	atcaggcatt	tttatacttt	gctgttagga	atgcaatgta	43860
gtacaaacct	ttgtagaagt	tgctttggaa	atgtctctca	gatgtacaaa	tgcattcaca	43920
ttttagattt	agcattcccg	ctttctgaga	cattattcaa	catgtatacg	tgtgcacata	43980
agatataata	ataacacgtt	tttccttcta	gtgtgttgct	tttaacctgt	agcttgaaaa	44040
aactctgctt	: tcattotttt	tttttgttt	ctgtcactgg	ctcagccctg	ctttcaattg	44100
tttatatgaa	ttgatgggtg	ttctggtctg	gttataatct	actttagttt	aagagtcact	44160
ttaaattata	a tgacatctga	tataagttgt	gttaggtaga	. aaattetgta	acciggaala	44220
ctgtaagtag	tttgtggcca	catttcatta	. gtattaaata	. ttatctctat	atatagtagg	44280
chatttaata	a ttcatatttt	atgatgcaat	. taagaaataa	tttttttctg	aagttggtag	44340
attottoata	a toccatoocc	cagtgtttct	caaagcatto	: tgggggatca	ctgtttgtca	44400
gaattagete	r cagtgattgt	tgaacatgca	. gggcctctgc	: tccactccac	gttgctacca	44460
ggacgetete	r caggtgagag	ctgggaagct	: gtagaagctg	, cagtgctaac	: aaatgctaca	44520
ggaattette	tagtcacctt	catgaggtct	: tatgttgagg	, agaggcagco	agtagtgtcc	44580
cttatectte	e cogttttato	gtgtaagttt	: cattttaagg	, gaggtataaa	tcaaagccca	44640
cctaggcati	t ctctcatagt	teactgette	: ttqtaatcat	: ggaagatgto	c attgcggcag	44700
agacgaaac	a ofotaottto	attactatto	, atttttttt	: aattatttt	ctgaagtggc	44760
tattataat	o taataaatto	r tgtgcttaag	, gacaaccttt	: ggtattctat	trgagtatig	44820
tgtatgatc	c tagttaagtt	: ttttctacca	a gtattttcat	: attacaacat	t atttacttc	44880
catttctat	t aatatttta	tatttaaagt	: atggaggcc	g ggcacagtgg	g ctcacgcgtg	44940
taatcccag	c attttgggat	gctgaggcgg	g gtggatcaca	a aggtcaggag	g ttctagacca	45000
gegtgaeca	a cacqqtqaaa	a tcccatctc1	t actaaaaata	a caaaaatta	g ccgggcacag	45060
tggtaggca	c ctgtaattc	agctactca	g gaggetgag	g taggagaat	c acttgaatcc	45120

gggaggcagc agttgcagtg agctaagatc gtgccactgg actctagcct ggctgacaga gcaagaatcc gcctaaaaaa aaagggatca gggaagaggg gattacagat aacccaaaga 45240 45300 agaaggaaaa atctccacaa gttcacctgt ccagcggtaa ccccaatttg gatattttcc tttaacaatt tggatatttt cctttaaatc ctctttttta taatgtctat atgttggaga gagtatgtgc ctttacgtat tttttaaaga tgagatttct gtgtgtgtct atatctcctg 45420 45480 ttcttcatat tttcttgtgt gttataaaca gctgtacatg tcagtatata tacttccgta 45540 acttttttt aaaggctata tagtgttcat tgatgtgatt taacagcagt tatctccccg 45600 gcttcatctt gttggaatgt gggtcctgtg tgttgccttc agagcaaatg gggcttggtt ttgcagcaag tagacctgtg acctgtacga atagttggaa gactttctct attacccaag 45660 cgtatcagta tactttagtg cctactagaa atttatgggt agaaaaacaa taatatctta 45720 gagtattttt tcctagattc cctaaggtgc tatagggtga tttttactca tgtaacatga actatgette aactaagata gtttttgeaa atgtggatat ataagtaett tattaaacet ataggaagta tttataccac ttatttcctc ccttcagtgt tagaacctcc taaatggcat 45900 ttgacattga actgctttcc actttgtcgc atgctcctct cattgtccct acctgggtcc 45960 tgaacettag ggaettgget gttatagece caccatgget acgetgggee ttggtegtet 46020 ctgagactta gtttcttcat cttacaagga gataataaca gcccctgcct gcgtagaatt 46080 gcagagatca aatgaaataa ttaacatact caaaagcatg ccgtaaacac attctgagca 46140 catgtacgtt ttaggaaaaa caaaaggacc catgcacatt tcggagtgct tttgtctcag 46200 cagcactgcc tottottoca aagctgacgt ottagtagag goodtgccac gtootgagca 46260 ctgtactcca cgaagcattc tatttctgac attcgaaatg cagtctgttc catcttcctt 46320 acaatetgta tgccagcaet tgaaataeeg ggtatetgea gtgttgaeea ggtgattaet 46380 taattatgga aatgttgagg tggagatcta gataattcag tgaaggcagg aaaattggtg 46440 46500 toggaatotg totttttatg tgtcagaaat agaaataaga tagggtgaga agtaatttgt ggctaaaaca ctataatagc taacacatag tgcatactgt gtgccaagca ctcctgtagg 46560 tgcttgaaat cttctattat tattatccct actttataga cttgcaccct taggcacaga 46620 gaggcggaca gttgtccaag gttaccccag aggtggagat ccaggctacc tgactccacc 46680 atgtgtgctc ttccctaggg cacagttgtg ctgctaaaaa tactttttaa gcagttcttt 46740 gattattcag atgatagtac tgtaggaaaa ttaagacaaa aataatgaaa aattaaaatc tttattttag tgttttgcac atgtattatt aaagccagtt tactcctgga agtgtgtaag 46860 aatacagggt atttttgatc acctaaatgc tgcatgttac taagagctcg acactgaagt 46920 caagaagagc agttgcagag agtacttagc aaaaacggga agtgtgtggg gttgaaggag 46980 47040 caaagacaag tottootogg acggtggagt gtagaattca toatttotoa gaacacgtot ttgaacgcat tttcaatttg aggccaaagg tctcagcctc ccactcggca tacctcccta 47100 ccttagtcag ctcttaaatc ttaggaatat ttctttgttc ttcaaggaac ttaaatatgt 47160 taacattctt acctgtccac agggagcccc ctacaaagaa gggagtttct agtctccgtt ctttcttgga ataaataata gcctcatacc ttgtgcaatc gaggctgaaa aagactgtct 47280 ccttttttca aataagcaag tcttagaaac tacagttgtt tacagggctc atggctattc 47340 cacagtaata attttggttc ttttaccaat tatataatat gttaaaatat ggcaagtatc 47400 aggaaagcaa ggagtggcaa tgattagaaa ccaatggcca agttagagag gaggggcaat 47460 tgctcccca agtttgttgt ggctgtgtag cagtcagtga cgagaagctg tgtgtcaggc 47520 gacaagcaaa gttgaggatt atcaggcgcc tgtgagtgcc cagctgtgtg ccaggtcagg 47580 aggtgccatc gtgagccaga ccagcttcct ctcggcccct gtggagctcg cagtctggtg 47640 gggaggcagc agtcaccatg gtgacaggtg acacactagg atggggctgg tggtggtagg 47700 catttgcggg tcccttcaga gaggtgagta tggacttaga ggaggctcca gcttcctatt 47760 cctgggctgt ctatagcact aaaagttgtc acatgaaaaa taacatttgg tactattgat 47820 ttaacttaat gacttatgta attgtagttg acttagaaat tataacatgc tcttctactt 47880 cagcttgaaa cccccaacca ccagtttata atccttttt tttaactttt gtttatttt 47940 48000 cctaaggaat ctgtactttt tcttcatttt acaacttttt ttgtcctgtt accttatttt catttttact ttatatgacc atgagttcta aaatagtaaa aaaaaagaat tatttttgtt 48060 ctttgttaga atttctctgc aaagaatgtc caaaaattca tattcacatt gatcgtatcg 48120 acaaaaaaga tgtcccagaa gaacaagaac atatgagaag atggctgcat gaacgtttcg 48180 aaatcaaaga taagtgagta acaacagttc cagcacttcc ggaacttcgg ttcaactaga 48240 tttcagtata gtcaacaatt tgaaaccaat gtaaatggtt atattgtctc aagaatacat 48300 tttataaatt caaatcaaat tttatgcatg tctgatcgtg ttttaaactt tacttgtaca 48360 aatcagtcta aaagaacttg ttacagtggg occatctact tgcattgata gtatttcttg 48420 gacaatacta cgtgataaca tagcaaatta aattaaaaac aacaacaaac acacaaaaaa 48480 actttccagt gtcagatgcc cggacctacc tgtcaggtca cataaagtgg tgttactgtg 48540 tgaggtctgg ctgttgggcc agtgtgcgca gaaaagcaag ggaggggtag aggactatgc 48600 ggacgtgcag gtggacatga tgctgttata tttgttggaa atagaagggg gcagttgaca 48660 gcgttatatc caaagtgtct tctgtggtta attatattca gaaattttag ccaattgttt 48720 tattetetaa atatgtaett tetgeteaag aaactateat tgttettett tteettgttt 48780 tacagtacag tgtttttaat taaccctcct gggttaactt taccaggtga aaatgattaa 48840 aagtgtaata ggttaacaat gaaactttaa gcttctattt ttcattgact cttaactgta 48900 catgatgtaa tgtattcagc gagccattca ggaccacttt ggcccatgga agaaatttaa 48960 aagtaagatc tacatgtatt gacatgaaaa tatgttctca gaaaaaagac taatgtattt 49020 aatgtcctac ttattttata agtatttaga atacctctgg acattttaaa acaatgatta 49080 ttgctagggt gtgtgattta taaagcaata gaagcgcttt ccctttctgt ttgtgtttta 49140 gattattata togggtatgt totgctatca taactttaca aatcttatgt aatatgggaa 49200 aatgagttaa ctatgetgtt tteettettt taeetgeett tetaattetg tgggaataaa 49260 ggcgtttttg agacagccca ggtgtagtga gcagtccata tccatggatt ccacattcat 49320 ggattccacc aagcacagac caaaaatact cagaaaaaaa gggggctggc tgtggtggct 49380 catgcatgta atcccagcac tttgggaggc taaggcaggc aaattgcttg agcccagaag 49440 ttcaagacag cctgggcaac atggcaaaac cctgtctcta cagaaaatac aaaaattagc 49500 caggegtgca cetgtagtee cagetactea ggaggeegag gtgegaggat caeetgagee tggaaggttg agactgcagt gagctatcat tgtgccaact ccagcetggt aacagagtgc 49620 cttttttcaa aaaaaaaaaa aaaaaaggat ttgggaggat atgcatatgt tatattcaaa 49680 tacatgccat tttattcata tatcagggac ttgagcatcc tttgatcttg gtctctgccg ggtatcctgg gaccagccc ctgtcgatac agagggaccg ctgtctaaga accgctggtc ctatctttga cttctggcgg aataggagct ccatgtaaaa aggaggagaa gctgcagcgg 49860 gttattagcc atttgtgagt caggtcactg taaaacttta tcaaaagttt aaaagacaaa 49920 aagcatecte ataaaatgee ttaaaaccae etgttgaaat attacatata caatteatgt 49980 50040 atactaatca tagagcatat taaagatatt ttagaagact agaaacttct attaaaccaa gtttctggat gtttccgtat tcatccttat tttccaggga cctgcataac ttttccagcg 50100 tgtaatagct acctgattga tattttttga attgaaatac tgaagtgact aaaatctaaa 50160 ctttttccat tctggccata ggatgcttat agaattttat gagtcaccag atccagaaag 50220 aagaaaaaga tttcctggga aaagtgttaa ttccaaatta agtatcaaga agactttacc atcaatgttg atcttaagtg gtttgactgc aggcatgctt atgaccgatg ctggaaggaa 50340 getgtatgtg aacacetgga tatatggaac cetacttgge tgeetgtggg ttactattaa 50400 agcatagaca agtagctgtc tccagacagt gggatgtgct acattgtcta tttttggcgg 50460 ctgcacatga catcaaattg tttcctgaat ttattaagga gtgtaaataa agccttgttg 50520 attgaagatt ggataataga atttgtgacg aaagctgata tgcaatggtc ttgggcaaac 50580 atacctggtt gtacaacttt agcatcgggg ctgctggaag ggtaaaagct aaatggagtt 50640 teteetgete tgtecattte etatgaacta atgacaactt gagaaggetg ggaggattgt 50700 gtattttgca agtcagatgg ctgcattttt gagcattaat ttgcagcgta tttcactttt 50760 tetgttattt teaatttatt acaacttgae ageteeaage tettattaet aaagtattta 50820 gtatcttgca gctagttaat atttcatctt ttgcttattt ctacaagtca gtgaaataaa 50880 ttgtatttag gaagtgtcag gatgttcaaa ggaaagggta aaaagtgttc atggggaaaa 50940 agctctgttt agcacatgat tttattgtat tgcgttatta gctgatttta ctcattttat 51000 atttgcaaaa taaatttcta atatttattg aaattgctta atttgcacac cctgtacaca 51060 cagaaaatgg tataaaatat gagaacgaag tttaaaaattg tgactctgat tcattatagc 51120 agaactttaa atttcccagc tttttgaaga tttaagctac gctattagta cttccctttg 51180 tetgtgccat aagtgettga aaacgttaag gttttetgtt ttgttttgtt tttttaatat 51240 caaaagagtc ggtgtgaacc ttggttggac cccaagttca caagattttt aaggtgatga 51300 gageetgeag acattetgee tagatttact agegtgtgee ttttgeetge ttetetttga 51360 tttcacagaa tattcattca gaagtcgcgt ttctgtagtg tggtggattc ccactgggct 51420 ctggtccttc ccttggatcc cgtcagtggt gctgctcagc ggcttgcacg tagacttgct 51480 51540 aggaagaaat gcagagccag cctgtgctgc ccactttcag agttgaactc tttaagccct tgtgagtggg cttcaccagc tactgcagag gcattttgca tttgtctgtg tcaagaagtt 51600 caccttctca agccagtgaa atacagactt aattcgtcat gactgaacga atttgtttat 51660 ttcccattag gtttagtgga gctacacatt aatatgtatc gccttagagc aagagctgtg 51720 ttccaggaac cagatcacga tttttagcca tggaacaata tatcccatgg gagaagacct 51780 ttcagtgtga actgttctat ttttgtgtta taatttaaac ttcgatttcc tcatagtcct 51840 ttaagttgac atttctgctt actgctactg gatttttgct gcagaaatat atcagtggcc 51900 cacattaaac ataccagttg gatcatgata agcaaaatga aagaaataat gattaaggga 51960 aaattaagtg actgtgttac actgcttctc ccatgccaga gaataaactc tttcaagcat 52020 catctttgaa gagtcgtgtg gtgtgaattg gtttgtgtac attagaatgt atgcacacat 52080 ccatggacac tcaggatata gttggcctaa taatcggggc atgggtaaaa cttatgaaaa 52140 tttcctcatg ctgaattgta attttctctt acctgtaaag taaaatttag atcaattcca 52200 tgtctttgtt aagtacaggg atttaatata ttttgaatat aatgggtatg ttctaaattt 52260 gaactttgag aggcaatact gttggaatta tgtggattct aactcatttt aacaaggtag 52320 cctgacctgc ataagatcac ttgaatgtta ggtttcatag aactatacta atcttctcac aaaaggtcta taaaatacag tcgttgaaaa aaattttgta tcaaaatgtt tggaaaatta 52440 gaagettete ettaacetgt attgatactg acttgaatta ttttetaaaa ttaagageeg 52500 tatacctacc tgtaagtctt ttcacatatc atttaaactt ttgtttgtat tattactgat 52560 ttacagctta gttattaatt tttctttata agaatgccgt cgatgtgcat gcttttatgt 52620 ttttcagaaa agggtgtgtt tggatgaaag taaaaaaaa aataaaatct ttcactgtct 52680 ctaatggctg tgctgtttaa cattttttga ccctaaaatt caccaacagt ctcccagtac 52740 ataaaatagg cttaatgact ggccctgcat tcttcacaat atttttccct aagctttgag 52800 caaagtttta aaaaaataca ctaaaataat caaaactgtt aagcagtata ttagtttggt 52860 tatataaatt catctgcaat ttataagatg catggccgat gttaatttgc ttggcaattc 52920 tgtaatcatt aagtgatctc agtgaaacat gtcaaatgcc ttaaattaac taagttggtg 52980 aataaaagtg ccgatctggc taactcttac accatacata ctgatagttt ttcatatgtt tcatttccat gtgattttta aaatttagag tggcaacaat tttgcttaat atgggttaca 53100 taagetttat ttttteettt gtteataatt atattetttg aataggtetg tgteaateaa 53160 gtgatctaac tagactgatc atagatagaa ggaaataagg ccaagttcaa gaccagcctg 53220 ggcaacatat cgagaacctg tctacaaaaa aattaaaaaa aattagccag gcatggtggc 53280 gtacactgag tagtttgtcc cagctactcg ggagggtgag gtgggaggat cgcttcagcc 53340 caggaggttg agattgcagt gagccatgga cataccactg cactacagcc taggtaacag 53400 cacgagaccc caactettag aaaatgaaaa ggaaatatag aaatataaaa tttgettatt 53460 53520 atagacacac agtaactccc agatatgtac cacaaaaaat gtgaaaagag agagaaatgt ctaccaaagc agtattttgt gtgtataatt gcaagcgcat agtaaaataa ttttaacctt 53580 aatttgtttt tagtagtgtt tagattgaag attgagtgaa atattttctt ggcagatatt 53640 ccgtatctgg tggaaagcta caatgcaatg tcgttgtagt tttgcatggc ttgctttata 53700 aacaagattt tttctccctc cttttgggcc agttttcatt acgagtaact cacacttttt 53760 gattaaagaa cttgaaatta cgttatcact tagtataatt gacattatat agagactatg 53820 taacatgcaa tcattagaat caaaattagt actttggtca aaatatttac aacattcaca 53880 tacttgtcaa atattcatgt aattaactga atttaaaacc ttcaactatt atgaagtgct 53940 cgtctgtaca atcgctaatt tactcagttt agagtagcta caactcttcg atactatcat caatatttga catcttttcc aatttgtgta tgaaaagtaa atctattcct gtagcaactg 54060 gggagtcata tatgaggtca aagacatata ccttgttatt ataatatgta tactataata 54120 atagctggtt atcctgagca ggggaaaagg ttatttttag gaaaaccact tcaaatagaa 54180 agctgaagta cttctaatat actgagggaa gtataatatg tggaacaaac tctcaacaaa 54240 atgtttattg atgttgatga aacagatcag tttttccatc cggattatta ttggttcatg 54300 attttatatg tgaatatgta agatatgttc tgcaatttta taaatgttca tgtctttttt 54360 54420 taaaaaaggt gctattgaaa ttctgtgtct ccagcaggca agaatacttg actaactctt tttgtctctt tatggtattt tcagaataaa gtctgacttg tgtttttgag attattggtg cctcattaat tcagcaataa aggaaaatat gcatctcaaa aattggtgat aaaaagttat 54540 ttcttgtata tgtgataaag tttacatgtt gtgtatatat gttgtattgc caaatacggc 54600 54660 tattaaatac tacgtcatat tttaaaggtt cagtttgtag tgatagtaaa caagcagtgc actaageete ttgegggeat cateteatet caetgteate acaaaceeca tgecacageg 54720 tagcttgacc actaaaagta atgcatctgc aagcatactg ccaggttttg gatagtttgt 54780 accaacagtt accttatcaa ggtaaatccc agactctaaa agagttggtg ctgtgtcact 54840 acatgcataa ctttaaataa atttcctgcc gggcgcggtg gctcacgcct gtaatcccag 54900 cagtttggga ggccgaggca agtggatcac ttgaggtcag gagtttgaga ccagcctggc 54960 caacgtggtg aaaccctgtc tctactaaaa atacaaaaat tagccaggcg tgtggtggca 55020 ggcacctgta atcccagcta cttgggagga tgaggcagga gaatcatttg aatcctgcag 55080 geggaggttg cagtgageca agatggegte attgeactee ageetgggeg acaagagega 55140 gactccgtat taaaaaaaaa aaaaaaaaaa aaaaaaaaatt cctctcctgt ttgagctttc 55200 ccttacctgt aaagagggga gaatatgtat ttacttcaaa gagttcaggg aaatgactct 55260 55320 cactagtttg agattctagg tataaaaata cattcttata taattttaac accaatgtga gagattatta ttcttgctaa accaattcag ttttatttgc tgtctaaaat gtgtgaataa 55380 gtaattgtcc attattttct gaagtgtttt ggaactcaac acatgattgt gaggaggatt 55440 tgttgctaaa catctttctg gttattcaag ctcgtgtata ctgtgctctg ttgagacatg 55500 cagagttact ttctgtctgg gtcacaggtc agttcttgat agttttcgga caattaacca 55560 gttttcattt gcccatgacc acctttattc tttttcctca actgcaccca tcttttataa ggtctttcag tttattgcag agaagatggt ggagaaaagc cggaattccc acccaccgct 55680 gccatcccca tgttttatca ttggctagag tggaaaatag cagtaactac tgtgagagat 55740 catttgttta tataatggaa acaaagatga ggaaagaacc tggcttagat cagagaactg 55800 atgtatttag attcttttt tttttttt taagacggag tgttgctctg ttgcccagac 55860 tggagtacag tggctcaatc tcggctcact gcaacctcca tttccctggt tcaagcaatt atcctgcctc agcctcccaa gtatttggga ttacaggcgt gttccaccac acctggctaa 55980 ttttttgtat ttttagtaga gacggggttt cgccatgttg gccaggctgg tctcgaaatc 56040 ctgacctcag atgatccacc cgccttggcc tcccaaagtg ctgggattac aggcgcgagc 56100

cacogogoct ggoccaatgt atttggattc ttaaagaaca ctttcaaatt aaatacagt gaaggaaga tagaagctaa gaattctgt gtcaactgt ttagcaaatg taagtagaag 25 (2009agagtg tgtcctggaa tgaatgaaa catcagtaaa ataccatacg tatgttataga tgttattgtt tecttgoctt ggttgatttg gtttattgt gtttattgt gattatt tcattadaga ttgtgattg gtttattag gtttgattgt gtttatagt gtttattgt gtttattgt caaggtgtc taaaaggca 26 (200) 2 (210) 2 (211) 2 (212) NA (213) Homo sapiens (400) 2 (210) 3 (212) DNA (213) Homo sapiens (220) 3 (212) DNA (213) Homo sapiens (220) 3 (212) DNA (213) Homo sapiens (220) 3 (212) DNA (213) Homo sapiens (220) 3 (212) DNA (213) Homo sapiens (220) 3 (212) DNA (213) Homo sapiens (220) 3 (212) DNA (213) Homo sapiens (221) PolyA_signal (222) 5180 . 5186 (223) ANTANA (400) 3 (223) ANTANA (400																
<pre><210 > 2 211 > 23 <212 > DNA</pre>	tgaagag ctgggag tgttatt attgtga tccttaa	aac taatg tagt to get to the terms to the terms to a que to the terms	agaad gtcct ccttg tggaa ttctg	ctaaa ggaa gccti attt gaggi	a gaa a tga t gg g gt t tg	attto aatga ttga! catc! ggat!	ctgt aata tttg tact taag	gtcate gtt aga gta	aaacı cagta ttacı aaatı gtgt	egt in a a a a a a a a a a a a a a a a a a	ttago ataco gaaa aaago caago	caaa cata taat aagt gtgt	tg ta cg ta tt to ta a tc ta	aagt atgt caat tagc aaaa	agaag tatga ataga tatct cggca	56280 56340 56400 56460
<pre>2212 > DNA 2213 > Homo sapiens</pre>			gcac.					3						-	_	
<pre><213 > Homo sapiens <400> 2 gtcgtccag cgcttggtag aag <210 > 3 <211> 5227 <212> DNA </pre> <pre><221> bolyA_signal <222> 51805186 <223> AATAAA <400> 3 ctgctgtccc tggtgctca cacgtactcc atg cgc tac ctg ctg ccc agc gtc</pre>																
Secretaria Sec			ລກໍຄາ													
Section Sect			apre	.15												
<pre></pre>			gctt	ggta	g aa	g										23
<pre></pre>																
<pre></pre>																
<pre></pre>		omo s	apie	ns												
<pre></pre>	_	olva	sian.	a l												
<400> 3 ctgctgtccc tggtgctcca cacgtactce atg cgc tac ctg ctg ccc agc gtc 54 Met Arg Tyr Leu Leu Pro Ser Val 1 5 gtg ctc ctg ggc acg ggc acg ggc acc acc tac gtg ttg gcc tgg ggg gtc tgg 102 Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp 10 20 cgg ctg ctc tcc gcc ttc ctc gcc ttc ctc gcc gc				~ ~												
Start Star																
Met Arg Tyr Leu Leu Pro Ser Val 1	<400> 3	.aaa +	aata	ctcc	a ca	cata	ctcc	ato	. כמכ	tac	cta	ata	ccc	agc	atc	54
Strict S	Cigcigi		.gg cg		u cu	cgca		Met	Arg	Tyr	Leu	Leu	Pro	Ser	Val	
Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp 10												-				100
10	gtg ctc	ctg	ggc	acg Thr	gcg Nla	CCC Pro	acc Thr	tac Tur	gtg Val	ttg Leu	gcc Ala	tgg Trp	ggg Glv	gtc Val	Tro	102
cgg ctg ctc tcc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac 150 Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp 30 25 30 35 40 gac cgg ctc tac tac tgc gtc tac cag agc atg gtg ctc ttc ttc ttc ttc gag 198 Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Phe Glu 55 aat tac acc ggg gtc cag ata ttg cta tat gga gat ttg cca aaa aat 246 Asn Tyr Thr Gly Val Gln Ile Leu Leu Tyr Gly Asp Leu Pro Lys Asn 60 70 aaa gaa aat ata ata tat tta gca aat cat caa agc aca gtt gac tgg 294 Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr Val Asp Trp 80 85 att gtt gct gac atc ttg gca atc atg ggc atc acg aat gtg ctg cag aat ggc cta gga cat gtg 342 Ile Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu Gly His Val 90 95 cgc tac gtg ctg aaa gaa gag tta aaa tgg ctg cca ttg tat ggg tgt Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys 115 120 105 110 115 tac ttt gct cag cat gga gaa atc tat gta aag cgc agt gcc aaa ttt Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe 125 130 aac gag aaa gag atg cga aca gag atg cga aca ggt tat ac gtg gac gag ga Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly 140 145 acc atg tat tct tat	10					15					20					
25	caa cto	ctc	tçc	gcc	ttc	ctg	ccc	gcc	cgc	ttc	tac	caa	gcg	ctg	gac	150
gac cgg ctc tac tgc gtc tac cag agc atg gtg ctc ttc ttc ttc ttc gag Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Phe Glu 45 50 55 aat tac acc ggg gtc cag ata ttg cta tat gga gat ttg cca aaa aat Asn Tyr Thr Gly Val Gln Ile Leu Leu Tyr Gly Asp Leu Pro Lys Asn 60 65 70 aaa gaa aat ata ata tat tta gca aat cat caa agc aca gtt gac tgg Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr Val Asp Trp 75 80 85 att gtt gct gac atc ttg gcc atc agg cag aat gcg cta gga cat gtg Ile Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu Gly His Val 90 95 cgc tac gtg ctg aaa gaa ggg tta aaa tgg ctg cca ttg tat ggg tgt Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys 105 110 tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe 125 130 120 acc ca agg aaa gag atg cga aac aag ttg cag agc tac gtg gac gag gac Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly 140 145 150 act cca atg tat ctt gtg att tt cca gaa ggt aca agg tat aat cca Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro 155 1 160 155 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 ggc ctt gca gta tta aaa cat gtg cta aca cca cca cag ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr	_	ı Leu	Ser	Ala		Leu	Pro	Ala	Arg		Tyr	GIn	Ala	ьеи	Asp 40	
Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Phe Glu	gac cg	ctc	tac	tac	gtc	tac	cag	agc	atg	gtg	ctc	ttc	ttc	ttc	gag	198
aat tac acc ggg gtc cag ata ttg cta tat gga gat ttg cca aaa aat 246 Asn Tyr Thr Gly Val Gln Ile Leu Leu Tyr Gly Asp Leu Pro Lys Asn 60 65 70 aaa gaa aat ata tat tta gca aat cat caa agc aca gtt gac tgg Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr Val Asp Trp 80 294 Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr Val Asp Trp 75 80 85 att gtt gct gac atc ttg gcc atc agg cag aat gcg cag aat gcg ctg cta gga cat gtg Ile Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu Gly His Val 90 95 cgc tac gtg ctg aaa gaa ggg tta aaa tgg ctg cca ttg tat ggg tgt Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys 105 110 tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aat ttt Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe 125 130 aac gag aaa gaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly 140 145 act cca atg tat ctt gtg att tt cca gaa ggt aca agg tat aat cca agg tat aat cca agg caa aca aca aca aca aca aca aca	Asp Arg	Leu	Tyr	Cys	Val	Tyr	Gln	Ser	Met	Val	Leu	Phe	Phe	Phe	Glu	
Asn Tyr Thr Gly Val Gln Ile Leu Leu Tyr Gly Asp Leu Pro Lys Asn 60 65 70 70 aaa gaa aat ata ata tat tta gca aat cat caa agc aca gtt gac tgg 294 Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr Val Asp Trp 80 85 att gtt gct gac atc ttg gcc atc agg cag aat gcg cta gga cat gtg 342 Ile Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu Gly His Val 90 95 100 cgc tac gtg ctg aaa gag ggt ta aaa tgg ctg cca ttg tat ggg tgt 390 Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys 110 120 tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt 438 Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe 125 130 135 aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga 486 Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly 140 145 150 act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aat cca Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro 155 160 165 165 165 170 gag caa aca aca aca gtc ctt ca gct agt cag gca ttt gct gcc caa cgt 582 Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Ala Gln Arg 170 175 180 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr Pro Arg Ile Lys Ala Thr	t				cad	ata	tta	cta		gga	gat	tta	cca		aat	246
aaa gaa aat ata ata tat tta gca aat cat caa agc aca gtt gac tgg Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr Val Asp Trp 75 80 810 825 826 826 837 841 838 839 842 842 843 844 845 845 845 845 845 845 845 845 845	Asn Tv	Thr	Gly	Val	Gln	Ile	Leu	Leu	Tyr	Gly	Asp	Leu	Pro	Lys	Asn	
Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr Val Asp Trp 75 att gtt gct gac atc ttg gcc atc agg cag aat gcg cta gga cat gtg Ile Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu Gly His Val 90 cgc tac gtg ctg aaa gaa ggg tta aaa tgg ctg cca ttg tat ggg tgt Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys 105 tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe 125 aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly 140 act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aat cca Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro 155 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr			60					65					70			204
att gtt gct gac atc ttg gcc atc agg cag aat gcg cta gga cat gtg Ile Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu Gly His Val 90 cgc tac gtg ctg aaa gaa ggg tta aaa tgg ctg cca ttg tat ggg tgt Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys 105 110 115 120 tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe 125 aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly 140 act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aat cca Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro 155 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr	aaa gaa	a aat	ata	ata Tlo	tat	tta	gca	Asn	His	Gln	Ser	Thr	Val	Asp	Trp	234
The Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu Gly His Val 90 cgc tac gtg ctg aaa gaa ggg tta aaa tgg ctg cca ttg tat ggg tgt Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys 105 110 tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe 125 130 aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly 140 act cca atg tat ctt gtg att ttc cca gaa ggt aca agg tat aat cca Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro 155 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr		75					80					85				
90 cgc tac gtg ctg aaa gaa ggg tta aaa tgg ctg cca ttg tat ggg tgt Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys 105 110 tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe 125 130 aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly 140 act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aat cca Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro 155 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr	att gt	gct	gac	atc	ttg	gcc	atc	agg	cag	aat	gcg	cta	gga	cat	gtg	342
cgc tac gtg ctg aaa gaa gaa gag tta aaa tgg ctg cca ttg tat ggg tgt 390 Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys 105 110 115 120 tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe 125 130 135 aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly 145 150 act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aaat cca 334 Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro 155 160 165 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 175 ggc ctt gca gta tta aaa cat gtg cta aca cca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr 200 630		l Ala	Asp	Ile	Leu		TTE	Arg	Gin	ASII	100	Leu	GTĀ	птѕ	Vai	
105 110 115 120 tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt 438 Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe 125 130 135 aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly 140 145 150 act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aat cca Tyr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro 155 160 165 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 582 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr 630	cac ta	c gtg	ctg	aaa	gaa	ggg	tta	aaa	tgg	ctg	cca	ttg	tat	ggg	tgt	390
tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe 125 130 135 aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly 140 145 150 act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aat cca Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro 155 160 165 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 175 180 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr	Arg Ty	r Val	Leu	Lys	Glu	Gly	Leu	Lys	Trp	Leu	Pro	Leu	Tyr	Gly	Cys	
Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe 125 aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly 140 act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aat cca Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro 155 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr		t act	cad	cat		gga	atc	tat	gta		cac	agt	acc	aaa		438
aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly 140 act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aat cca Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro 155 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr	Tyr Ph	e Ala	Gln	His	Gly	Gly	Ile	Tyr	Val	Lys	Arg	Ser	Ala	Lys	Phe	
Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly 140 act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aat cca Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro 155 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr				125					130					135		106
act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aat cca 534 Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro 155 160 165 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 175 180 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr	aac ga	g aaa	gag	atg	cga	aac Aen	aag	Leu	Gln	agc	Tac Tvr	gtg Val	Asp	gca	gga Glv	400
Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro 155 160 165 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt 582 Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 175 180 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr	ASII GI	u nys		Mec	AL 9	nan	Lys		01	501	-,-				1	
gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 175 180 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr	act co	a atg	tat	ctt	gtg	att	ttt	cca	gaa	ggt	aca	agg	tat	aat	cca	534
gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 175 180 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr	Thr Pr			Leu	Val	Ile			Glu	GIY	Thr		ıyr	ASN	Pro	
Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg 170 175 180 ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr	gag ca	a aca	aaa	gtc	ctt	tca	gct	agt	cag	gca	ttt	gct	gcc	caa	cgt	582
ggc ctt gca gta tta aaa cat gtg cta aca cca cga ata aag gca act Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr	Glu Gl	n Thr	Lys	Val	Leu	Ser	Āla	Ser	Gln	Ala	Phe	Ala	Ala	Gln	Arg	
Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr	17	0		++-	227		at a	cta	aca	CCA		ata	aac	aca	act	630
- 105	Gly Le	u Ala	. yta . Val	Leu	Lys	His	Val	Leu	Thr	Pro	Arg	Ile	Lys	Ala	Thr	0.50
	_		!													

cac His	gtt Val	gct Ala	ttt Phe	gat Asp 205	tgc Cys	atg Met	aag Lys	aat Asn	tat Tyr 210	tta Leu	gat Asp	gca Ala	att Ile	tat Tyr 215	gat Asp	678
gtt Val	acg Thr	gtg Val	gtt Val 220	tat	gaa Glu	ggg	aaa Lys	gac Asp 225	gat	gga Gly	ggg ggg	cag Gln	cga Arg 230	aga	gag Glu	726
tca Ser	ccg Pro	acc Thr 235	atg	acg Thr	gaa Glu	ttt Phe	ctc Leu 240	tgc Cys	aaa Lys	gaa Glu	tgt Cys	cca Pro 245	aaa Lys	att Ile	cat His	774
Ile	His 250	Ile	Asp	Arg	Ile	Asp 255	Lys	Lys	Asp	gtc Val	Pro 260	Glu	Glu	Gln	Glu	822
His 265	Met	Arg	Arg	Trp	Leu 270	His	Glu	Arg	Phe	gaa Glu 275	Ile	Lys	Asp	Lys	Met 280	870
Leu	Ile	Glu	Phe	Tyr 285	Glu	Ser	Pro	Asp	Pro 290	gaa Glu	Arg	Arg	Lys	Arg 295	Phe	918
Pro	Gly	Lys	Ser 300	Val	Asn	Ser	Lys	Leu 305	Ser	atc Ile	Lys	Lys	Thr 310	Leu	Pro	966
Ser	Met	Leu 315	Ile	Leu	Ser	Gly	Leu 320	Thr	Ala	Gly	Met	Leu 325	Met	Thr	Asp	1014
Ala	Gly 330	Arg	Lys	Leu	Tyr	Val 335	Asn	Thr	Trp	ata Ile	Tyr 340	Gly	Thr	Leu	Leu	1062
Gly 345	Cys	Leu	Trp	Val	Thr 350	Ile	Lys	Ala	*	aca						1112
agt	ggga	tgt	gcta	catt	gt c	tatt	tttg	g cg	gctg	caca	tga	catc	aaa	ttgt	ttcctg	1172 1232
aat	ttat	taa	ggag	tgta	aa t	aaag	cctt	g tt	gatt	gaag	att	ggat	aat	agaa	tttgtg	1232
acg	aaag	ctg	atat	gcaa	tg g	ctaa	gggc	a aa	ttct	ceta	ctc	tata	cat	ttcc	gcatcg tatgaa	1352
ggg	etge	cgg	aagg	gtaa	aa y aa c	taaa	acyy	t to	tata	tttt	gca	agto	aga	taac	tgcatt	1412
ttt	acya	att	aatt	taca	ac a	tatt	tcac	t tt	ttct	gtta	ttt	tcaa	ttt	atta	caactt	1472
gac	aget	CCA	aget	ctta	tt a	ctaa	agta	t tt	agta	tctt	gca	gcta	gtt	aata	tttcat	1532
ctt	ttac	tta	tttc	taca	ag t	cagt	gaaa	t aa	attg	tatt	tag	gaag	tgt	cagg	atgttc	1592
aaa	agaa	agg	gtaa	aaaq	ta t	tcat	gggg	a aa	aago	tctg	ttt	agca	cat	gatt	ttattg	1652
tat	taca	rtta	ttag	rctga	tt t	tact	catt	t ta	ıtatt	tgca	aaa	taaa	ittt	ctaa	tattta	1712
ttg	jaaat	tgc	ttaa	itttg	cac	acco	tgta	c ac	acag	aaaa	tgg	rtata	laaa	tatg	agaacg	1772
aag	ıttta	aaa	ttgt	gact	.ct g	atto	atta	it aç	ıcaga	actt	taa	attt	ccc	agct	ttttga	1832
aga	ittta	agc	taco	ctat	ta g	tact	tccc	t tt	gtct	gtgc	cat	aagt	gct	tgaa	aacgtt	1892 1952
aag	gttt	tct	gttt	tgtt	בב ב	וכככנ		la ta	iccaa	laaga	gu	ggu	tet	acct	tggttg	2012
gad	CCCC	agt	tcac	aaga	ice t	actt	ctct	t to	jayaç tətt	caca	rae	tatt	cat	tcac	agattt aagtcg	2072
cat	-ttct	ata	atat	-aatc	raa t	tccc	acto	ia ac	tete	ratco	tto	cctt	gga	tccc	gtcagt	2132
aai	acto	ictc	acco	racti	ac a	cata	gact	t a	tago	raaga	aat	gcac	agc	cago	ctgtgc	2192
ta	cca	ttt	caga	agtto	raa o	tctt	taaa	ag co	ctto	tgag	r tgg	gctt	cac	cago	tactgc	2252
aga	agge	attt	taca	attt	rtc t	gtgt	caaç	ga ag	gttca	acctt	cto	caago	cag	tgaa	atacag	2312
act	ttaai	tca	tcat	tgact	ga a	acgaa	atttg	gt t1	tattt	ccca	i tta	aggti	tag	tgga	igctaca	2372
ca	ttaat	tatq	tate	cgcct	ta g	gagca	agag	go to	gtgtt	ccag	gaa	accag	gatc	acga	ittttta	2432
ac	catg	aac	aata	atato	CC 8	atggg	gagaa	ag a	cctt	cagt	gto	gaaci	tgtt	ctat	ttttgt	2492
gt	tata	attt	aaa	cttc	gat t	tcci	cata	ag to	cctt	aagt	tga	acati	LECE	gcti	actgct	2552 2612
ac	tgga	cttt	tgc	rgca	gaa a	atata	atcag	gt g	auss. Acccs	aCdtl	. ddi	traci	tata	tta	ggatcat cactgct	2672
ga	taag	caaa	atg	aaagi	add 1	acto	yalli ttta:	aa y	gyadi	atctt	. ay	aana	atca	tate	ggtgtga	2732
a+	taat	ttat	ota	uyaa catt:	aga a	atat	atac	ac a	catc	cato	aca	acto	agga	tata	agttggc	2792
ct	aata	atco	aaa	cato	aat a	aaaa	ctta	tg a	aaat	ttcci	ca	tgct	gaat	tgt	aattttc	2852
tc	ttac	ctat	aaa	gtaa	aat	ttag	atca	at t	ccat	gtcti	t tg	ttaa	gtac	agg	gatttaa	2912
ta	tatt	ttga	ata	taat	ggg	tatg	ttct	aa a	tttg	aact	t tg	agag	gcaa	tac	tgttgga	2972

WO 99/32644

```
attatgtgga ttctaactca ttttaacaag gtagcctgac ctgcataaga tcacttgaat
                                                                  3032
gttaggtttc atagaactat actaatcttc tcacaaaagg tctataaaat acagtcgttg
                                                                   3092
aaaaaaattt tgtatcaaaa tgtttggaaa attagaagct tctccttaac ctgtattgat
                                                                   3152
actgacttga attattttct aaaattaaga googtatacc tacctgtaag tottttcaca
                                                                   3212
tatcatttaa acttttgttt gtattattac tgatttacag cttagttatt aatttttctt
                                                                   3272
tataagaatg ccgtcgatgt gcatgctttt atgtttttca gaaaagggtg tgtttggatg
                                                                   3332
3392
ttgaccctaa aattcaccaa cagtctccca gtacataaaa taggcttaat gactggccct
                                                                   3452
gcattcttca caatattttt ccctaagctt tgagcaaagt tttaaaaaaaa tacactaaaa
                                                                   3512
taatcaaaac tgttaagcag tatattagtt tggttatata aattcatctg caatttataa
gatgcatggc cgatgttaat ttgcttggca attctgtaat cattaagtga tctcagtgaa
                                                                   3632
acatgtcaaa tgccttaaat taactaagtt ggtgaataaa agtgccgatc tggctaactc
                                                                   3692
ttacaccata catactgata gtttttcata tgtttcattt ccatgtgatt tttaaaattt
                                                                   3752
agagtggcaa caattttgct taatatgggt tacataagct ttattttttc ctttgttcat
                                                                   3812
                                                                   3872
aattatattc tttgaatagg tctgtgtcaa tcaagtgatc taactagact gatcatagat
agaaggaaat aaggccaagt tcaagaccag cctgggcaac atatcgagaa cctgtctaca
                                                                   3932
aaaaaattaa aaaaaattag ccaggcatgg tggcgtacac tgagtagttt gtcccagcta
                                                                   3992
ctcgggaggg tgaggtggga ggatcgcttc agcccaggag gttgagattg cagtgagcca
                                                                   4052
tggacatacc actgcactac agcctaggta acagcacgag accccaactc ttagaaaatg
                                                                   4112
aaaaggaaat atagaaatat aaaatttgct tattatagac acacagtaac tcccagatat
                                                                   4172
                                                                   4232
gtaccacaaa aaatgtgaaa agagagagaa atgtctacca aagcagtatt ttgtgtgtat
                                                                   4292
aattgcaagc gcatagtaaa ataattttaa ccttaatttg tttttagtag tgtttagatt
gaagattgag tgaaatattt tcttggcaga tattccgtat ctggtggaaa gctacaatgc
                                                                   4352
aatgtcgttg tagttttgca tggcttgctt tataaacaag attttttctc cctccttttg
                                                                   4412
ggccagtttt cattacgagt aactcacact ttttgattaa agaacttgaa attacgttat
                                                                   4472
                                                                   4532
cacttagtat aattgacatt atatagagac tatgtaacat gcaatcatta gaatcaaaat
tagtactttg gtcaaaatat ttacaacatt cacatacttg tcaaatattc atgtaattaa
                                                                   4592
ctgaatttaa aaccttcaac tattatgaag tgctcgtctg tacaatcgct aatttactca
                                                                   4652
gtttagagta gctacaactc ttcgatacta tcatcaatat ttgacatctt ttccaatttg
                                                                   4712
tgtatgaaaa gtaaatctat tcctgtagca actggggagt catatatgag gtcaaagaca
                                                                   4772
tataccttgt tattataata tgtatactat aataatagct ggttatcctg agcaggggaa
                                                                   4832
aaggttattt ttaggaaaac cacttcaaat agaaagctga agtacttcta atatactgag
                                                                   4892
ggaagtataa tatgtggaac aaactctcaa caaaatgttt attgatgttg atgaaacaga
                                                                   4952
                                                                   5012
tcagtttttc catccggatt attattggtt catgatttta tatgtgaata tgtaagatat
gttctgcaat tttataaatg ttcatgtctt tttttaaaaa aggtgctatc gaaattctgt
                                                                   5072
gtctccagca ggcaagaata cttgactaac tctttttgtc tctttatggt attttcagaa
                                                                   5132
                                                                   5192
taaagtotga ottgtgtttt tgagattatt ggtgootcat taattoagca ataaaggaaa
                                                                   5227
atatgcattt caaaaanaaa aaaaaaaaaa aaaaa
 <210> 4
 <211> 353
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> HELIX
 <222> 1..33
 <223> Rao and Argos identification method, potential helix
 <221> HELIX
 <222> 4..20
 <223> Klein, Kanehisa and DeLisi identification method, potential helix
 <221> HELIX
 <222> 4..24
 <223> Eisenberg, Schwarz, Komarony, Wall identification method,
 potential helix
 <221> MYRISTATE
 <222> 12..16
 <223> Prosite match
 <221> HELIX
 <222> 50..70
 <223> Eisenberg, Schwarz, Komarony, Wall identification method,
 potential helix
 <221> CARBOHYD
```

WO 99/32644 PCT/IB98/02133

```
<222> 57..59
<223> Prosite match
<221> HELIX
<222> 76..96
<223> Eisenberg, Schwarz, Komarony, Wall identification method,
potential helix
<221> PHOSPHORYLATION
<222> 78
<223> potential Tyrosine kinase site, Prosite match
<221> PHOSPHORYLATION
<222> 84
<223> potential caseine kinase II site, Prosite match
<221> SITE
<222> 94..115
<223> potential Leucine zipper site, Prosite match
<221> MYRISTATE
<222> 119..123
<223> potential site, Prosite match
<221> PHOSPHORYLATION
<222> 133
<223> potential protein kinase C, Prosite match
<221> PHOSPHORYLATION
<222> 147
<223> potential caseine kinase II site, Prosite match
<221> PHOSPHORYLATION
<222> 194
<223> potential protein kinase C, Prosite match <221> PHOSPHORYLATION
<222> 215
<223> potential Tyrosine kinase site, Prosite match
<221> SULFATATION
<222> 221
<223> Prosite match
<221> PHOSPHORYLATION
<222> 233
<223> potential cAMP and cGMP dependant protein kinase site, Prosite
match
<221> PHOSPHORYLATION
<222> 235
<223> potential caseine kinase II site, Prosite match
<221> PHOSPHORYLATION
<222> 306
<223> potential protein kinase C, Prosite match <221> HELIX
<222> 310..330
 <223> Eisenberg, Schwarz, Komarony, Wall identification method,
 potential helix
 <221> MYRISTATE
 <222> 319..323
 <223> Prosite match
 <221> MYRISTATE
 <222> 323..327
 <223> Prosite match
 <221> AMIDATION
 <222> 329
 <223> Prosite match
 <221> HELIX
 <222> 333..353
 <223> Eisenberg, Schwarz, Komarony, Wall identification method,
 potential helix
 <221> MYRISTATE
```

```
<222> 341..345
<223> Prosite match
<221> PHOSPHORYLATION
<222> 350
<223> potential protein kinase C, Prosite match
<400> 4
Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
                              25
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
                           40
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
                       55
Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala
                70
Asn His Gln Ser Thr Val Asp Trp Ile Val Ala Asp Ile Leu Ala Ile
                                 90
              85
Arg Gln Asn Ala Leu Gly His Val Arg Tyr Val Leu Lys Glu Gly Leu
                                                 110
                             105
Lys Trp Leu Pro Leu Tyr Gly Cys Tyr Phe Ala Gln His Gly Gly Ile
                                              125
                           120
  115
Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg Asn Lys
                                        140
                    135
Leu Gln Ser Tyr Val Asp Ala Gly Thr Pro Met Tyr Leu Val Ile Phe
                 150
                                     155
Pro Glu Gly Thr Arg Tyr Asn Pro Glu Gln Thr Lys Val Leu Ser Ala
                                  170
                                                     175
               165
Ser Gln Ala Phe Ala Ala Gln Arg Gly Leu Ala Val Leu Lys His Val
                                                190
                               185
           180
 Leu Thr Pro Arg Ile Lys Ala Thr His Val Ala Phe Asp Cys Met Lys
                                              205
        195
                          200
 Asn Tyr Leu Asp Ala Ile Tyr Asp Val Thr Val Val Tyr Glu Gly Lys
                                220
                      215
 Asp Asp Gly Gly Gln Arg Arg Glu Ser Pro Thr Met Thr Glu Phe Leu
            230
                                       235
 Cys Lys Glu Cys Pro Lys Ile His Ile His Ile Asp Arg Ile Asp Lys
                                   250
               245
 Lys Asp Val Pro Glu Glu Glu His Met Arg Arg Trp Leu His Glu
                               265
            260
 Arg Phe Glu Ile Lys Asp Lys Met Leu Ile Glu Phe Tyr Glu Ser Pro
                                              285
                          280
       275
 Asp Pro Glu Arg Arg Lys Arg Phe Pro Gly Lys Ser Val Asn Ser Lys
                                          300
                       295
 Leu Ser Ile Lys Lys Thr Leu Pro Ser Met Leu Ile Leu Ser Gly Leu
                                       315
                   310
 Thr Ala Gly Met Leu Met Thr Asp Ala Gly Arg Lys Leu Tyr Val Asn
                                             335
                                  330
               325
 Thr Trp Ile Tyr Gly Thr Leu Leu Gly Cys Leu Trp Val Thr Ile Lys
                              345
 <210> 5
 <211> 364
 <212> PRT
 <213> Homo sapiens
 <400> 5
 Met Leu Leu Ser Leu Val Leu His Thr Tyr Ser Met Arg Tyr Leu Leu
                                   10
               5
 Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp
                                25
            2.0
 Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln
```

```
40
       35
Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe
 50 55
                            60
Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu Leu Tyr Gly Asp Leu
                                    75
                  70
Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr
                               90
              85
Val Asp Trp Ile Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu
                            105
          100
Gly His Val Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu
                                          125
                         120
Tyr Gly Cys Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser
                                        140
                     135
Ala Lys Phe Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val
                 150
                                  155
Asp Ala Gly Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg
                                                 175
                              170
              165
Tyr Asn Pro Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala
                                       190
                            185
         180
Ala Gln Arg Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg Ile
                      200
                                            205
      195
Lys Ala Thr His Val Ala Phe Asp Cys Met Lys Asn Tyr Leu Asp Ala
                      215
                                       220
Ile Tyr Asp Val Thr Val Val Tyr Glu Gly Lys Asp Asp Gly Gly Gln
                                    235
                  230
Arg Arg Glu Ser Pro Thr Met Thr Glu Phe Leu Cys Lys Glu Cys Pro
                                 250
                                                 255
              245
Lys Ile His Ile His Ile Asp Arg Ile Asp Lys Lys Asp Val Pro Glu
                            265
                                               270
          260
Glu Gln Glu His Met Arg Arg Trp Leu His Glu Arg Phe Glu Ile Lys
                          280
                                           285
      275
Asp Lys Met Leu Ile Glu Phe Tyr Glu Ser Pro Asp Pro Glu Arg Arg
                                     300
                     295
  290
Lys Arg Phe Pro Gly Lys Ser Val Asn Ser Lys Leu Ser Ile Lys Lys
                          315 320
305 310
Thr Leu Pro Ser Met Leu Ile Leu Ser Gly Leu Thr Ala Gly Met Leu
       325 330
Met Thr Asp Ala Gly Arg Lys Leu Tyr Val Asn Thr Trp Ile Tyr Gly
                         345
  340
Thr Leu Leu Gly Cys Leu Trp Val Thr Ile Lys Ala
                         360
<210> 6
<211> 26
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..26
<223> primer oligonucleotide GC1.5p.1
<400> 6
                                                                26
ctgtccctgg tgctccacac gtactc
<210> 7
<211> 26
<212> DNA
<213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..26
 <223> primer oligonucleotide GC1.5p.2
 <400> 7
                                                                 26
 tggtgctcca cacgtactcc atgcgc
```

```
<210> 8
<211> 27
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..27
<223> primer oligonucleotide pg15RACV196 <400> 8
                                                                       27
caatatctgg accccggtgt aattctc
<210> 9
<211> 34
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..34
<223> primer oligonucleotide GC1.3p
<400> 9
                                                                       34
cttqcctqct ggagacacag aatttcgata gcac
<210> 10
<211> 24
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..24
<223> primer oligonucleotide PGRT32
<400> 10
                                                                       24
tttttttt ttttttttg aaat
<210> 11
<211> 6
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 160..165
<223> box2 from SEQID4, present in AF003136, P33333, P26647, U89336,
U56417, AB005623.
<400> 11
Phe Pro Glu Gly Thr Arg
<210> 12
 <211> 6
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> 129..134
 <223> box2 from Z72511
 <400> 12
 Phe Pro Glu Gly Thr Asp
 <210> 13
 <211> 6
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> 223..228
 <223> box2 from P38226, Z49770
```

```
<400> 13
Phe Pro Glu Gly Thr Asn
              5
<210> 14
<211> 6
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 90..95
<223> box2 from Z49860 and Z29518
<400> 14
Phe Val Glu Gly Thr Arg
<210> 15
<211> 9
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 211..219
<223> box3 from SEQID4, present in AF003136
Leu Asp Ala Ile Tyr Asp Val Thr Val
1
<210> 16
<211> 9
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 204..212
<223> box3 from Z72511
<400> 16
Val Glu Tyr Ile Tyr Asp Ile Thr Ile
               5
1
<210> 17
<211> 9
<212> PRT
<213> Homo sapiens
<220>
<221> SITE <222> 271..279
<223> box3 from P38226
 <400> 17
 Ile Glu Ser Leu Tyr Asp Ile Thr Ile
                5
 <210> 18
 <211> 9
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> 265..273
 <223> box3 from Z49770
 <400> 18
 Leu Asp Ala Ile Tyr Asp Val Thr Ile
                 5
 <210> 19
 <211> 9
 <212> PRT
```

```
<213> Homo sapiens
<220>
<221> SITE
<222> 138..146
<223> box3 fromZ49860
<400> 19
Val Pro Ala Ile Tyr Asp Met Thr Val
<210> 20
<211> 9
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 218..226
<223> box3 from Z29518
<400> 20
Val Pro Ala Ile Tyr Asp Thr Thr Val
                 5
<210> 21
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-123
 <221> allele
 <222> 24
 <223> polymorphic base C
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-123.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-123.mis2
 <400> 21
                                                                         47
 tttctcatcc tcacacctca ctgcgcccct cctgaaccca ctccttt
 <210> 22
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-26
 <221> allele
  <222> 24
  <223> polymorphic base G
  <221> primer_bind
  <222> 1..23
  <223> potential microsequencing oligo 4-26.mis1
  <221> primer_bind
  <222> 25..47
  <223> complement potential microsequencing oligo 4-26.mis2
  <400> 22
                                                                         47
  ccctgtnaga cacgtcctgt atcgttgttg agatgggaaa gtgcatc
  <210> 23
  <211> 47
  <212> DNA
  <213> Homo Sapiens
```

PCT/IB98/02133 WO 99/32644 27

```
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-14
<221> allele
<222> 24
<223> polymorphic base T
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-14.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-14.mis2
<400> 23
                                                                        47
gcagggagca gaccagacat gatttgttct agtctagctg attcata
<210> 24
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-77, extracted from SEQ ID1 12057 12103
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-77.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-77.mis2
                                                                        47
gctgttcaga ctaaacttgg agactacagt cagtcagaga acttgct
<210> 25
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-217, extracted from SEQ ID1 34469 34515
<221> allele
 <222> 24
 <223> polymorphic base C
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-217.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-217.mis2
 <400> 25
                                                                         47
 atatagtica cgttatgttc atacttaatt gttgcatttt gtttgcc
 <210> 26
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-67, extracted from SEQ ID1 51612 51658
```

WO 99/32644 PCT/IB98/02133 28

```
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-67.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-67.mis2
<400> 26
                                                                       47
gccagtgaaa tacagactta attcgtcatg actgaacgaa tttgttt
<210> 27
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-213
<221> allele
<222> 24
<223> polymorphic base T
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-213.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-213.mis2
<400> 27
                                                                       47
ccttagcatt caagcccctg agctctggtg ttgtccaccc ctggggg
<210> 28
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-221
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-221.mis1
<221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-221.mis2
 <400> 28
                                                                        47
 agcttgagaa accagaaaag ccaaaaggag gctcctacca catgggt
 <210> 29
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-135
 <221> allele
 <222> 24
 <223> polymorphic base A
 <221> primer_bind
```

```
<222> 1..23
<223> potential microsequencing oligo 99-135.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-135.mis2
<400> 29
                                                                       47
agtcactata tctatgttta atgaagatag aaagagatgc agaaatg
<210> 30
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-123, variant version of SEQ ID21
<221> allele
<222> 24
<223> base T ; C in SEQ ID21
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-123.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-123.mis2
<400> 30
                                                                        47
tttctcatcc tcacacctca ctgtgcccct cctgaaccca ctccttt
<210> 31
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
 <222> 1..47
<223> polymorphic fragment 4-26, variant version of SEQ ID22
 <221> allele
 <222> 24
 <223> base A ; G in SEQ ID22
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-26.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-26.mis2
 <400> 31
                                                                         47
 ccctgtnaga cacgtcctgt atcattgttg agatgggaaa gtgcatc
 <210> 32
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-14, variant version of SEQ ID23
  <221> allele
  <222> 24
  <223> base C ; T in SEQ ID23
  <221> primer_bind
  <222> 1..23
  <223> potential microsequencing oligo 4-14.mis1
  <221> primer_bind
  <222> 25..47
```

```
<223> complement potential microsequencing oligo 4-14.mis2
gcagggagca gaccagacat gatctgttct agtctagctg attcata
                                                                       47
<210> 33
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-77, variant version of SEQ ID24
<221> allele
<222> 24
<223> base G ; C in SEQ ID24
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-77.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-77.mis2
<400> 33
                                                                        47
gctgttcaga ctaaacttgg agagtacagt cagtcagaga acttgct
<210> 34
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-217, variant version of SEQ ID25
<221> allele
<222> 24
<223> base T ; C in SEQ ID25
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-217.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-217.mis2
 <400> 34
                                                                        47
 atatagttca cgttatgttc atatttaatt gttgcatttt gtttgcc
 <210> 35
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-67, variant version of SEQ ID26
 <221> allele
 <222> 24
 <223> base T ; C in SEQ ID26
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-67.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-67.mis2
                                                                         47
  gccagtgeaa tacagactta atttgtcatg actgaacgaa tttgttt
  <210> 36
```

31

```
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-213, variant version of SEQ ID27
<221> allele
<222> 24
<223> base C ; T in SEQ ID27
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-213.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-213.mis2
<400> 36
                                                                        47
cettageatt caageceetg agecetggtg ttgtecaeee etggggg
<210> 37
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-221, variant version of SEQ ID28
<221> allele
<222> 24
<223> base C ; A in SEQ ID28
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-221.mis1
<221> primer_bind <222> 25..47
<223> complement potential microsequencing oligo 99-221.mis2
                                                                         47
 agcttgagaa accagaaaag ccacaaggag gctcctacca catgggt
 <210> 38
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-135, variant version of SEQ ID29
 <221> allele
 <222> 24
 <223> base G ; A in SEQ ID29
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-135.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-135.mis2
                                                                          47
 agtcactata tctatgttta atggagatag aaagagatgc agaaatg
 <210> 39
  <211> 18
  <212> DNA
  <213> Homo Sapiens
  <220>
```

MICHAEL AND ORDERAND 1 -

PCT/IB98/02133

```
<221> primer_bind
<222> 1..18
<223> upstream amplification primer 99-123-PU
<400> 39
                                                                       18
aaagccagga ctagaagg
<210> 40
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer 4-26-PU
<400> 40
                                                                        18
tacagccctg taagacac
<210> 41
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer 4-14-PU
<400> 41
                                                                        18
tctaacctct catccaac
<210> 42
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer 4-77-PU, extracted from SEQ ID1 11930
 11947
 <400> 42
                                                                         18
 tgttgattta caggcggc
 <210> 43
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> upstream amplification primer 99-217-PU, extracted from SEQ ID1 34216
 34234
 <400> 43
                                                                         19
 ggtgggaatt tactatatg
 <210> 44
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
  <221> primer_bind
  <223> upstream amplification primer 4-67-PU, extracted from SEQ ID1 51596
  51613
  <400> 44
                                                                          18
  aagttcacct tctcaagc
  <210> 45
  <211> 20
  <212> DNA
```

<213> Homo Sapiens <220>	
<220> <221> primer_bind	
<222> 120	
<223> upstream amplification primer 99-213-PU	
<400> 45	20
atactggcag cgtgtgcttc	20
<210> 46 <211> 19	
<211> 19 <212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> upstream amplification primer 99-221-PU	
<400> 46	19
ccctttttct tcactgttc <210> 47	
<211> 18	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 118 <223> upstream amplification primer 99-135-PU	
<223 upstream amplification primer 33 13 13 (400> 47)	
tggaagttgt tattgccc	18
<210> 48	
<211> 18	
<212> DNA	
<213> Homo Sapiens	
<220>	
<pre><221> primer_bind c222> 1 18</pre>	
<pre><222> 118 <223> downstream amplification primer 99-123-RP</pre>	
<400> 48	
tattcagaaa ggagtggg	18
<210> 49	
<211> 18	
<212> DNA	
<213> Homo Sapiens	
<220> <221> primer_bind	
<222> 118	
<223> downstream amplification primer 4-26-RP	
<400> 49	10
tgaggactgc taggaaag	18
<210> 50	
<211> 20 <212> DNA	
<2112> DNA <213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 120	
<223> downstream amplification primer 4-14-RP	
<400> 50	20
gactgtatcc tttgatgcac	20
<210> 51 <211> 20	
<211> 20 <212> DNA	
<213> Homo Sapiens	

PCT/IB98/02133 34

```
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer 4-77-RP, extracted from SEQ ID1 12339
58 complement
<400> 51
                                                                        20
ggaaaggtac tcattcatag
<210> 52
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer 99-217-RP, extracted from SEQ ID1 34625
34645 complement
<400> 52
                                                                        21
gtttattttg tgtgagcttt g
<210> 53
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer 4-67-RP, extracted from SEQ ID1 51996
520
15 complement
<400> 53
                                                                         20
tgaaagagtt tattctctgg
<210> 54
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer 99-213-RP
 <400> 54
                                                                         21
 ttattgcccc acatgcttga g
 <210> 55
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> downstream amplification primer 99-221-RP
 <400> 55
                                                                         19
 tcattcgtct ggctaggtc
 <210> 56
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> downstream amplification primer 99-135-RP
 <400> 56
                                                                          18
 aaacacctcc cattgtyc
```

```
<210> 57
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1482
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1482.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1482.mis2
                                                                        47
agtgaagtct gagggggaaa aatcaaccct atagagggaa ggatctg
<210> 58
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-73, extracted from SED ID1 13657 13703
<221> allele
<222> 24
<223> polymorphic base C in PG1 (13680) SEQ ID1
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-73.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-73.mis2
<400> 58
                                                                        47
gttttcctta tgatgttaca tggcttattt ttaaaggtaa tgaaaac
<210> 59
<211> 47
<212> DNA
<213> Homo Sapiens
 <220>
<221> allele
 <222> 1..47
 <223> polymorphic fragment 4-65, extracted from SEQ ID1 51448 51494
 <221> allele
 <222> 24
 <223> polymorphic base T in PG1 (51471) SEQ ID1
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-65.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-65.mis2
 <400> 59
                                                                         47
 ggtgctgctc agcggcttgc acgtagactt gctaggaaga aatgcag
 <210> 60
 <211> 47
 <212> DNA
 <213> Hor o Sapiens
```

```
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1482, variant version of SEQ ID57
<221> allele
<222> 24
<223> base A ; C in SEQ ID57
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1482.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1482.mis2
<400> 60
                                                                        47
agtgaagtct gagggggaaa aataaaccct atagagggaa ggatctg
<210> 61
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-73, variant version of SEQ ID58
<221> allele
<222> 24
<223> base G ; C in SEQ ID58
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-73.mis1
<221> primer_bind
<222> 25..47
 <223> complement potential microsequencing oligo 4-73.mis2
 <400> 61
                                                                         47
 gttttcctta tgatgttaca tgggttattt ttaaaggtaa tgaaaac
 <210> 62
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-65, variant version of SEQ ID59
 <221> allele
 <222> 24
 <223> base C ; T in SEQ ID59
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-65.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-65.mis2
 <400> 62
                                                                         47
 ggtgctgctc agcggcttgc acgcagactt gctaggaaga aatgcag
  <210> 63
  <211> 21
  <212> DNA
  <213> Homo Sapiens
  <220>
  <221> primer_bind
  <222> 1..21
  <223> upstream amplification primer 99-1482-PU
```

```
<400> 63
                                                                       21
atcaaatcag tgaagtctga g
<210> 64
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer 4-73-PU, extracted from SEQ ID1 13547
13564
<400> 64
                                                                        18
atcgctggaa cattctgg
<210> 65
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> upstream amplification primer 4-65-PU, extracted from SEQ ID1 51149
51168
<400> 65
                                                                        20
gatttaagct acgctattag
<210> 66
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer 99-1482-RP
<400> 66
                                                                        20
acaaatctat ataaggctgg
<210> 67
<211> 20
<212> DNA
<213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..20
 <223> downstream amplification primer 4-73-RP, extracted from SEQ ID1 13962
 13981 complement
 <400> 67
                                                                        20
 ctcttggtta aacagcagtg
 <210> 68
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> downstream amplification primer 4-65-RP, extracted from SEQ ID1 51482
 51499 complement
 <400> 68
                                                                         18
 tggctctgca tttcttcc
 <210> 69
 <211> 5226
 <212> DNA
 <213> Homo sapiens
 <400> 69
```

			ggtg					Met 1	Arg	Tyr	Leu	Leu 5	ı Pro	Ser	Val	54
Val	Leu 10	Leu	ggc Gly	Thr	Ala	Pro 15	Thr	Tyr	Val	Leu	Ala 20	Trp	Gly	Val	Trp	102
Arg 25	ctg Leu	Leu	tcc Ser	Ala	Phe 30	Leu	Pro	Ala	Arg	Phe 35	Tyr	Gln	Ala	Leu	Asp 40	150
Asp	Arg	Leu	tac Tyr	Cys 45	Val	Tyr	Gln'	Ser	Met 50	Val	Leu	Phe	Phe	Phe 55	Glu	198
Asn	Tyr	Thr	ggg Gly 60	Val	Gln	Ile	Leu	Leu 65	Tyr	Gly	Asp	Leu	Pro 70	Lys	Asn	246
Lys	Glu	Asn 75	ata Ile	Ile	Tyr	Leu	Ala 80	Asn	His	Gln	Ser	Thr 85	Val	Asp	Trp	294
Ile	Val 90	Ala	gac Asp	Ile	Leu	Ala 95	Ile	Arg	Gln	Asn	Ala 100	Leu	Gly	His	Val	342
Arg 105	Tyr	Val	ctg Leu	Lys	Glu 110	Gly	Leu	Lys	Trp	Leu 115	Pro	Leu	Tyr	GIY	Cys 120	390
tac Tyr	ttt Phe	Ala	cag Gln	His 125	Gly	Gly	Ile	Туr	Val 130	Lys	Arg	Ser	Ala	Lys 135	Pne	438
Asn	Glu	Lys	gag Glu 140	Met	Arg	Asn	Lys	Leu 145	Gln	Ser	Tyr	Val	Asp 150	Ala	GIY	486
Thr	Pro	Met 155	tat Tyr	Leu	Val	Ile	Phe 160	Pro	Glu	Gly	Thr	Arg 165	Tyr	Asn	Pro	534
Glu	Gln 170	Thr	aaa Lys	Val	Leu	Ser 175	Ala	Ser	Gln	Ala	Phe 180	Ala	Ala	Gln	Arg	582
Gly 185	Leu	Ala	gta Val	Leu	Lys 190	His	Val	Leu	Thr	Pro 195	Arg	Ile	Lys	Ala	Thr 200	630
His	. Val	Ala	ttt a Phe	Asp 205	Суз	Met	Lys	Asn	Tyr 210	Leu	Asp	Ala	Ile	Tyr 215	Asp	678
Val	Thr	Va]	g gtt L Val 220	. Tyr	Glu	Gly	. Tàs	Asp 225	Asp	Gly	Gly	*				727
cac	cgac	cat	gacg	gaat	tt c	tctg	rcaaa	g aa	tgtc	caaa	aat	tcat	att	caca	ttgatc	787
gta	itcga	caa	aaaa	igato	itc c	caga	agaa	.c aa	gaac	atat	gag	gaaga	tgg	ctgc	atgaac	847 907
gtt	tcga	aat	caaa	igata	ag a	tgct	tata	g aa		atga	t gtc	caeca	igat raan	actt	aaagaa	967
gaa	aaaag	jact	tcct	ggga	laa c	tas	taai	ים מכ	aaal	ttat	g cac	caat	act	ggaa	taccat ggaagc	1027
ca:	acgtt	.gat	Cacc	ayu	yyu t	.cyac	raacc	e ta	ictic	racto	, gat	atac	artt	acta	ttaaag	1087
cat	tagar	.yaa	tage	ctato	tc o	agag	agto	ig de	itgto	ctac	att	gtct	att	tttc	gegget	1147
aca	acato	raca	tcaa	aatto	gtt t	cctg	gaatt	t at	taaç	ggagt	: gta	aaata	aaag	cctt	gttgat	1207
ta	aagat	taa	ataa	ataga	aat t	tgtg	gacga	a ag	gctga	atato	g caa	atggt	cctt	gggc	caaacat	1267
ac	ctaal	ttat	acaa	actti	aq o	catco	gggg	t go	tgga	aaggg	y taa	aaago	ctaa	atgg	gagtttc	1327
to	ctgc	tctg	tcca	attt	cct a	tgaa	actaa	at ga	acaa	cttga	a gaa	aggct	ggg	agga	attgtgt	1387 1447
at	tttg	caag	tcag	gatg	gct (gcati	- C C C C	ja go	cacta	aditi	L gca	29091 att=/	ctas	ants	etttttc atttagt	1507
rg	ctat'	LTTC	taati	ttar'	tat i	ttca	rotti	ty C	cttai	tttet	t aca	aaat	cagt	gaaa	ataaatt	1567
at	attt:	agga	agt	atca	gga '	tatt	caaa	gg a	aagge	gtaa	a aa	gtgt	tcat	ggg	gaaaaag	1627
ct	ctgt	ttag	cac	atga	ttt	tatt	gtati	tg c	gtta	ttag	c tg	attt	tact	cati	tttatat	1687

ttgcaaaata aatttctaat atttattgaa attgcttaat ttgcacaccc tgtacacaca 1747 gaaaatggta taaaatatga gaacgaagtt taaaattgtg actctgattc attatagcag 1807 aactttaaat ttcccagctt tttgaagatt taagctacgc tattagtact tccctttgtc 1867 1927 aaagagtcgg tgtgaacctt ggttggaccc caagttcaca agatttttaa ggtgatgaga 1987 gcctgcagac attctgccta gatttactag cgtgtgcctt ttgcctgctt ctctttgatt 2047 tcacagaata ttcattcaga agtcgcgttt ctgtagtgtg gtggattccc actgggctct 2107 ggtccttccc ttggatcccg tcagtggtgc tgctcagcgg cttgcacgta gacttgctag 2167 2227 gaagaaatgc agagccagcc tgtgctgccc actttcagag ttgaactctt taagcccttg tgagtgggct tcaccagcta ctgcagaggc attttgcatt tgtctgtgtc aagaagttca 2287 cetteteaag ceagtgaaat acagaettaa ttegteatga etgaacgaat ttgtttattt 2347 cccattaggt ttagtggagc tacacattaa tatgtatcgc cttagagcaa gagctgtgtt 2407 ccaggaacca gatcacgatt tttagccatg gaacaatata tcccatggga gaagaccttt 2467 cagtgtgaac tgttctattt ttgtgttata atttaaactt cgatttcctc atagtccttt 2527 aagttgacat trotgottac tgctactgga tttttgctgc agaaatatat cagtggccca 2587 cattaaacat accagttgga tcatgataag caaaatgaaa gaaataatga ttaagggaaa 2647 attaagtgac tgtgttacac tgcttctccc atgccagaga ataaactctt tcaagcatca 2707 tetttgaaga gtegtgtgt gtgaattggt ttgtgtacat tagaatgtat gcacacatee 2767 atggacactc aggatatagt tggcctaata atcggggcat gggtaaaact tatgaaaatt 2827 tecteatget gaattgtaat tttetettae etgtaaagta aaatttagat caatteeatg 2887 tetttgttaa gtacagggat ttaatatatt ttgaatataa tgggtargtt etaaatttga 2947 actttgagag gcaatactgt tggaattatg tggattctaa ctcattttaa caaggtagcc 3007 tgacctgcat aagatcactt gaatgttagg tttcatagaa ctatactaat cttctcacaa 3067 aaggtctata aaatacagtc gttgaaaaaa attttgtatc aaaatgtttg gaaaattaga 3127 agetteteet taacetgtat tgatactgae ttgaattatt ttetaaaatt aagageegta 3187 tacctacctg taagtctttt cacatatcat ttaaactttt gtttgtatta ttactgattt 3247 acagettagt tattaatttt tetttataag aatgeegteg atgtgeatge ttttatgttt 3307 3367 aatggctgtg ctgtttaaca ttttttgacc ctaaaattca ccaacagtct cccagtacat 3427 aaaataggct taatgactgg ccctgcattc ttcacaatat ttttccctaa gctttgagca 3487 aagttttaaa aaaatacact aaaataatca aaactgttaa gcagtatatt agtttggtta 3547 tataaattca totgoaattt ataagatgca tggccgatgt taatttgctt ggcaattctg 3607 taatcattaa gtgatctcag tgaaacatgt caaatgcctt aaattaacta agttggtgaa 3667 taaaagtgcc gatctggcta actcttacac catacatact gatagttttt catatgtttc 3727 3787 atttccatgt gatttttaaa atttagagtg gcaacaattt tgcttaatat gggttacata agetttattt ttteetttgt teataattat attetttgaa taggtetgtg teaateaagt 3847 gatctaacta gactgatcat agatagaagg aaataaggcc aagttcaaga ccagcctggg 3907 caacatatcg agaacctgtc tacaaaaaaa ttaaaaaaaa ttagccaggc atggtggcgt 3967 acactgagta gtttgtccca gctactcggg agggtgaggt gggaggatcg cttcagccca 4027 ggaggttgag attgcagtga gccatggaca taccactgca ctacagccta ggtaacagca 4087 cgagacccca actcttagaa aatgaaaagg aaatatagaa atataaaatt tgcttattat 4147 agacacacag taactcccag atatgtacca caaaaaatgt gaaaagagag agaaatgtct 4207 accaaagcag tattttgtgt gtataattgc aagcgcatag taaaataatt ttaaccttaa 4267 tttgttttta gtagtgttta gattgaagat tgagtgaaat attttcttgg cagatattcc 4327 gtatctggtg gaaagctaca atgcaatgtc gttgtagttt tgcatggctt gctttataaa 4387 caagattttt teteeetet tttgggeeag tttteattae gagtaaetea caetttttga 4447 ttaaagaact tgaaattacg ttatcactta gtataattga cattatatag agactatgta 4507 acatgcaatc attagaatca aaattagtac tttggtcaaa atatttacaa cattcacata 4567 cttgtcaaat attcatgtaa ttaactgaat ttaaaacctt caactattat gaagtgctcg 4627 tctgtacaat cgctaattta ctcagtttag agtagctaca actcttcgat actatcatca 4687 atatttgaca tettttecaa tttgtgtatg aaaagtaaat etatteetgt agcaactggg 4747 gagtcatata tgaggtcaaa gacatatacc ttgttattat aatatgtata ctataataat 4807 agetggttat cetgageagg ggaaaaggtt atttttagga aaaccaette aaatagaaag 4867 ctgaagtact totaatatac tgagggaagt ataatatgtg gaacaaactc tcaacaaaat 4927 gtttattgat gttgatgaaa cagatcagtt tttccatccg gattattatt ggttcatgat 4987 tttatatgtg aatatgtaag atatgttctg caattttata aatgttcatg tctttttta 5047 aaaaaggtgc tatcgaaatt ctgtgtctcc agcaggcaag aatacttgac taactctttt 5107 tgtctcttta tggtattttc agaataaagt ctgacttgtg tttttgagat tattggtgcc 5167 tcattaattc agcaataaag gaaaatatgc atttcaaaaa naaaaaaaaa aaaaaaaaa 5226 <210> 70

<211> 228

```
<212> PRT
<213> Homo sapiens
<400> 70
Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
                                    10
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
                                25
           20
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
                                                45
                            40
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
                                            60
                        55
Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala
                                        75
                    70
Asn His Gln Ser Thr Val Asp Trp Ile Val Ala Asp Ile Leu Ala Ile
                85
                                    90
Arg Gln Asn Ala Leu Gly His Val Arg Tyr Val Leu Lys Glu Gly Leu
                                                    110
                                105
Lys Trp Leu Pro Leu Tyr Gly Cys Tyr Phe Ala Gln His Gly Gly Ile
                                                125
                            120
        115
Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg Asn Lys
                                           140
    130
                        135
Leu Gln Ser Tyr Val Asp Ala Gly Thr Pro Met Tyr Leu Val Ile Phe
                   150
                                        155
Pro Glu Gly Thr Arg Tyr Asn Pro Glu Gln Thr Lys Val Leu Ser Ala
                                                         175
                                    170
                165
Ser Gln Ala Phe Ala Ala Gln Arg Gly Leu Ala Val Leu Lys His Val
                                                    190
                                 185
            180
Leu Thr Pro Arg Ile Lys Ala Thr His Val Ala Phe Asp Cys Met Lys
                            200
                                                205
        195
Asn Tyr Leu Asp Ala Ile Tyr Asp Val Thr Val Val Tyr Glu Gly Lys
                                            220
                        215
   210
Asp Asp Gly Gly
225
<210> 71
<211> 158
<212> DNA
<213> Homo sapiens
<400> 71
gccttgcagt attaaaacat gtgctaacac cacgaataaa ggcaactcac gttgcttttg
attgcatgaa gaattattta gatgcaattt atgatgttac ggtggtttat gaagggaaag
                                                                      120
                                                                      158
acgatggagg gtagcgaaga gagtcaccga ccatgacg
 <210> 72
 <211> 1381
 <212> DNA
 <213> Mus musculus
 <220>
 <221> misc_binding
 <222> 608..629
 <223> amplification primer g34292.pu
 <221> misc_binding
 <222> 740..758
 <223> amplification primer g34292.rp
 <400> 72
 gagecgagag gatgetgetg tecetggtge tecacaegta etet atg ege tae etg
                                                                        56
                                                  Met Arg Tyr Leu
 ctc ccc age gtc ctg ttg ctg ggc tcg gcg ccc acc tac ctg ctg gcc
                                                                       104
 Leu Pro Ser Val Leu Leu Gly Ser Ala Pro Thr Tyr Leu Leu Ala
                                         15
                     10
 tgg acg ctg tgg cgg gtg ctc tcc gcg ctg atg ccc gcc cgc ctg tac
                                                                       152
 Trp Thr Leu Trp Arg Val Leu Ser Ala Le. Met Pro Ala Arg Leu Tyr
```

				25					30					35		
cag Gln	cgc Arg	gtg Val	Asp	gac	cgg Arg	ctt Leu	tac Tyr	tgc Cys	gtc	tac Tyr	cag Gln	aac Asn	atg Met 50	gtg Val	ctc Leu	200
ttc Phe	ttc Phe	ttc Phe	40 gag Glu	aac Asn	tac Tyr	acc Thr	Gly	45 gtc Val	cag Gln	ata Ile	ttg Leu	Leu	tat	gga Gly	gat Asp	248
ttg Leu	cca Pro	55 aaa Lys	aat Asn	aaa Lys	gaa Glu	Asn	60 gta Val	ata Ile	tat Tyr	cta Leu	Ala	55 aac Asn	cat His	caa Gln	agc Ser	296
Thr	70 gtt Val	gac Asp	tgg Trp	att Ile	gtt Val 90	75 gcg Ala	gac Asp	atg Met	ctg Leu	gct Ala 95	80 gcc Ala	aga Arg	cag Gln	gat Asp	gcc Ala 100	344
85 cta Leu	gga Gly	cat His	gtg Val	cgc Arg 105	tac Tyr	gta Val	ctg Leu	aaa Lys	gac Asp 110	aag Lys	tta Leu	aaa Lys	tgg Trp	ctt Leu 115	ccg Pro	392
ctg Leu	tat Tyr	ggg Gly	ttc Phe 120	tac	ttt Phe	gct Ala	cag Gln	cat His 125	gga Gly	gga Gly	att Ile	tat Tyr	gta Val 130	aaa Lys	cga Arg	440
agt Ser	gcc Ala	aaa Lys 135	EFE	aat Asn	gat Asp	aaa Lys	gaa Glu 140	atg	aga Arg	agc Ser	aag Lys	ctg Leu 145	cag Gln	agc Ser	tat Tyr	488
gtg Val	aac Asn 150	aca	gga Gly	aca Thr	ccg Pro	atg Met 155	tat	ctt Leu	gtg Val	att Ile	ttc Phe 160	cca Pro	gag Glu	gga Gly	aca Thr	536
Arg	tat	aat Asn	gca Ala	aca Thr	tac Tyr 170	aca	aaa Lys	ctc Leu	ctt Leu	tca Ser 175	gcc Ala	agt Ser	cag Gln	gca Ala	ttt Phe 180	584
165 gct Ala	gct Ala	cag Gln	cgg Arg	ggc Gly 185	ctt Leu	gca Ala	gta Val	tta Leu	aaa Lys 190	cac	gta Val	ctg Leu	aca Thr	cca Pro 195	aga Arg	632
ata Ile	aag Lys	gcc Ala	act Thr	cac His	gtt Val	gct Ala	ttt Phe	gat Asp 205	tct	atg Met	aag Lys	agt Ser	cat His 210	tta Leu	gat Asp	680
gca Ala	att Ile	tat Tyr 215	gat Asp	ato	aca Thr	gtg Val	gtt Val 220	tat Tyr	gaa Glu	ggg	aat Asn	gag Glu 225	. ьуѕ	ggt Gly	tca Ser	728
gga Gly	aaa Lys 230	tac Tyr	tea	aat Asn	cca Pro	cca Pro 235	tcc	atg	act Thr	gag Glu	ttt Phe 240	Lev	tgc Cys	aaa Lys	cag Gln	776
Cys	cca Pro		t ctt	cat His	att : Ile : 250	cac His	ttt	gat Asp	cgt Arg	ata Ile 255	gac Asp	aga	aat Asn	gaa Glu	gtt Val 260	824
245 cca Pro	rac	g gaa ı Glu	a caa ı Glr	a gaa n Glu 269	a cac ı His	ato	aaa Lys	aag Lys	tgg Trp	ctt Lev	: cat	gag Glu	g cgc	ttt p Phe 275	gag Glu	872
ata Ile	a aaa e Lys	a gat s Asj	t agg	g tto g Lei	r ctc	ata ı Ile	gaç Glu	tto Phe 285	tat Tyr	gat	tca Ser	a cca	a gat o Asp 29(Pro	gaa Glu	920
aga Arg	a aga	a aa g As: 29	c aaa n Ly:	a tti	t cct e Pro	o Gl7	g aaa / Lys 300	a agt s Ser	gtt	cai L Hi:	t tco	age r Are	g Let	a agt ı Sei	gtg Val	968
aa; Ly:	g aa s Ly 31	g ac s Th	+ ++	a cc u Pr	t tca	a gto r Val	g ttg l Le	g ato u Ile	ttg Lei	ı Gl	g ag y Se: 32	r Le	g act	t gcg r Ala	g gtc a Val	1016
at Me 32	g ct t Le	σ at	g ac t Th	g ga r Gl	g tc u Se 33	c gga r Gl	a ag	g aaa	a ct	g ta u Ty 33	c at	g gg	c acc	c tgg	g ttg p Leu 340	1064
ta	t. aa	a ac y Th	c ct ir Le	c ct u Le 34	t gg u Gl	c tg	c ct s Le	g tg u Tr	g tt p Ph 35	t gt e Va	t at	t aa e Ly	a gc s Al	a ta a * 35		1109

```
gcaagtagca ggctgcagtc acagtctctt attgatggct acacattgta tcacattgtt
                                                                  1169
1229
taageettga tgatngnnen ennnnnnnn nenantenng ngaccacage caacatgeat
                                                                  1289
ttgatttggg gcaaacacat gtggcttttc aggtgctggg gttgctggag acatggaagc
                                                                  1349
                                                                  1381
taagtggagt ttatgctgtt ttttttttt tt
<210> 73
<211> 15766
<212> DNA
<213> Mus musculus
<220>
<221> exon
<222> 52..121
<223> exon2
<221> exon
<222> 682..797
<223> exon3
<221> exon
<222> 2628..2717
<223> exon4
<221> exon
<222> 7834..7924
<223> exon5
<221> exon
<222> 9804..9965
<223> exon6
<221> exon
<222> 11404..11527
<223> exon7
<221> exon
<222> 13539..14035
<223> exon8
<221> misc_feature
<222> 13762..13764
<223> stop CDS
 <221> polyA_signal
 <222> 13835..13839
 <223> AATAAA : potential
 <400> 73
 ttttttttt ttaattgtca aagtcatgat tctttttgtt ttctctttta gatattgcta
                                                                    60
 tatggagatt tgccaaaaaa taaagaaaat gtaatatatc tagcgaatca tcaaagcaca
 ggtttgtatt tcatttgatg aaatttgggt ttttctagaa atggtaaatg agcattaata
                                                                    180
 tgtacacaca catacacaca aacacacata tgtacacaca catatgtttt aaagacagga
                                                                    240
 tttcatgtga cccagaatgg cctcatactc tctgagtagc tgagaatgat tttaagcttg
                                                                    300
 tgacacacct gccttcatct ccaaggtaca ggaattgcag gtgctttctt tgnnnnnnn
                                                                    360
 nnttttttt tttgagtttt ggggagggg tatatttttt aatgtgtctg tagttggctt
                                                                    420
 tgttttaagc attttaatca tactttattt ttaaaaaaac taaaagcttt tttaaggcta
                                                                    480
 ggtcttgcta tgtggcccta gtgttcctgg gacttgctct gtacaccggg ttgactctga
                                                                    540
 geetgtgege ettetgeete tgeeteeata gttagattet caggacatgt tacaaagaet
                                                                    600
 gtgctgtgaa gatgagtttt tgttcctggg agggaaggtt ggagctgact tgtgaggtac
                                                                    660
 tgacttgggt ctgccttaca gttgactgga ttgttgegga catgctggct gccagacagg
                                                                    720
 atgecetagg acatgtgege tacgtactga aagacaagtt aaaatggett eegetgtatg
                                                                    780
 ggttctactt tgctcaggta aactttgtct ttgccctttt atttcaaact taacaccatt
                                                                    840
 taatgaaact atatctgatt tttttgttta tgtgtttgtt ttatggtacc cgtgattgaa
                                                                    900
 catggggtca tatgtgtgct actgagtgac agccttagtt cagacatttt ttaaagcgac
                                                                    960
 ttttactagt atttttattt agaattctat atgtgtgcac atgcatatgt gtgcttgtgt
                                                                   1020
 gcacacgtgg atgcatgtga ggtcgaagga caattttcag tacaagtgtg agtgtcactt
                                                                   1080
 tttaggcacc ttccactctt attttgagac agtctcctag acctttgctg agttgcccag
                                                                   1140
 gctagccggc cagtgagccc tgggcatcta ccggtctctg cctccttacc tttacttagg
                                                                   1200
 ttacaagtgt gtgctgctac gcccagctgt ttactagatt ctagggatcc aaatgtgggt
                                                                    1260
 cctcgtaact tgtgagacaa gtactttcca aactgagcca cctccctagc tcttcttcac
                                                                    1320
 ggttcctgat ggtgtgtgtc tagatggctg gttgtccgta tatttaagtc cagtagcaga 1380
```

aatacaaata cctaggagtc caatagaaag ctacaagtgc agaattgaca atcggtaatg 1440 ttcggaaatt gattcaaaag tagttagtga gtgacagaca ggagctaaaa gcagactctg 1500 agctcagagt gtgaagtgtg gagaaatgtg ttttctcaca gttctgaagg ctgaaagtct 1560 ccccaaggtc aggatgtggg tggtactgct gtctcccaan cacccacctc tttggattat 1620 agactgcagc cttctccctg tgttctgagc cggcctttcc cacatgtgga catccttggt 1680 gggtgttcca ccagcagggc ctcagctagt gcccttattt cacttaactg taatgatttt 1740 cttaaagacc ctgtctccat acacagtcac tgtggaagct gaagcttcaa tgtaagagtt 1800 aaggggggag ggggaaattt agtccataat ggtgtcacac caatctctgt agctgagtcc 1860 atgattcagt totttaaagg ctctgagtgt agacattatc ttaattattt tgcccattta 1920 tgtattatct ttaatttatt ttatgtaact gaatgeetgt gtatatatgt ttetggttee 1980 2040 tagtocatat tttaattoot taaaggatgg aggtgtagac ttttgtottt ttaattttot atcottccct cotggcctcc tgtggcctct tttacgtatt tattatttt aatttatttt 2100 atgtgtttga gtgttttgta tctatgcatt cctggggccc atggaggtca gaaaaacaca 2160 ttaggtggcc tacaactgag ttatgggtgg tttgtgacca tggggtgctg ggacttgatc 2220 accagtetet gagaagaete gtgtetgetg ageettetet ceagteetgg gagtgtgggat 2280 attttaagga tacttttaat tgacttggtg aatgacagta gaaaatcaat gagttaggat 2340 ccatcggaaa aagcttttga actaaatctt ttaaagagaa aatattttaa gtgctaacaa 2400 aattaaatgt gtattttcca tgatgcagtt ttacttgggc tctgtagaaa taggattttc 2460 aggtacatat tgtatatata gttggcaata tttaaatact aactgtcgct tgagttctga 2520 aatgtagttt tatgtttttt actcattagg agtacagttg ccttaataac tacggagatt 2580 agttattaaa gaataattgc tottottttt tottttotgt gtaccagcat ggaggaattt 2640 atgtaaaacg aagtgccaaa tttaatgata aagaaatgag aagcaagctg cagagctatg 2700 tgaacgcagg aacaccggta agtgcgcccg cttttattcc tcaaggcagg ttaagaagtt 2760 aagttettaa gteattttga aaatatatta eeceatgtgg ageaatggaa etggtteggg 2820 2880 gttttgttga gataagctgt cctctggccg tgaggtaaga ttgctgcagg tgattgtaag gtttctcctg agtaacagtc agcatgggct cgggacgggc aagggcaggc cttagtgtgc 2940 agaggatgga gctcactgaa gccccaaaga gttagtcttc acatgagatt cagttctaga 3000 agaagttaaa ttgctttctt tctgtgtaaa tttggatttt tattgtagaa attaaagttt 3060 gttttctttt aaaacaaaca caaacccaga gcaaagagtc tcctagtgaa gagtcattcc 3120 gtgtcagtat tttacacaac tgtttttctg taaaggggga aaaagaattc aaatcttctc 3180 tttcaagaat gctgactgct gccaactgcc tctccccgtg gcccctctct gtatagacag 3240 gcatagctat ggtgaggact tgggcggctc ttgtctttct cctctctctg cttctctacc 3300 3360 ctttctctcg tgccctccac ttaccaggcc ctgggaagct acacaccagg caacagtgac 3420 cagggeeteg geetgggett egaccaatta etagageaga aacageagea getgeagtgt 3480 tgttttgtgc tgtgcactgt attaggttgt gttttcatca cctttgggtt ttgtgatgtt ttgatgaagt cctggtacca ttctagtttt tacattctgg gtagatagag tttattcaag 3540 gtctcaaggc atatgaatgg aagagctcct ctttacagcc attcgtgtag catgcataac 3600 tgctcttctg tattctctct agtgtctttt tttttgtgtg tgaatctgat gtcttgttat 3660 tcacctacaa tgtggagtaa tggtcataaa catataaagt acttatgcct ttatctgcca 3720 3780 aaaaaaattg aataccagcc tgttatagtg gcatatgcct gtgttcctag cactcaggag 3840 acaaaggcag aagtgtgaga acttcagact catactcagc tatatacaag accccaaatt 3900 3960 tgtgctagat tctgcagtac agccatgagt gtccccatct tagagggaga tcgctcatcc ttgtgctgtt ctttaagtct taccctgcaa cccactgtaa gtacactctt gctcacagtc 4020 ctttagaatc tcacactctt tctctttaca gacaccatgt cattgcccac tttattattt 4080 4140 atctgatgtc tacaaagatt atgaaagaga aacttgtatg cattctgtgt aaagtacttg acacaaataa tagtattcaa gaatgacttc ttaaatgaac actgaatgaa tagtttgttc 4200 taatttttt gatcaacaaa tcaaaaaata tttagattaa atatctaaga tacaaagcat 4260 aataccacat gaatcattaa agtgagtaat caatcttata agtgactgac cctaaaactc 4320 atagacatta ataattgctt tcattgctta gatataaact ttattgatta atacgttctc 4380 atgaaagtgg ttcttggaag gttctggaaa cgaaaatatt tttcttactg cttttttctt 4440 ctagtaactg attgaatttt tctgcagttc cataaagcat ctggtcaatt gctattatcc 4500 aatatgagga tatataacaa agtattgatt tttaaatttg gcggtgataa gacaagactg 4560 ggcgtgtgaa tgagggggtc tctgtttctt gtcccttctc ttgggttctt ttccttttgt 4620 tggtttgcct tctccagctg ctatgtgatg ggttctgatt atcttattat atcttatttt 4680 gttatttttc attgttatct cttagaagcc aacatgttat atcacctcca ctcccaccat 4740 taggtgtnct cacaaatacc ccaagctaaa caaccacatc atgtcatgtt nctgtatact 4800 tccataagtg ttgttaactc tactgactct tgtgagcagg cccaattggc tttatcccta 4860 gctgggtgac ctgggttcct ccccaacacc ataccgtcca tcaaactgag tccttttcca 4920 agcacacac agatactgct catctgagga ctcttctcat ccacctaagg actgcctgct 4980 cctcggcaga aagggcctct agtcccatac ccttacgccc tcaccaatgc cttaggaaca 5040 tgtgctcaat gcccctgtgg gtcatttccg tttacagtag ggaaatttgc ctgataactt 5100 gcagcacacc tataaagagg ccttgcttgc tctcatattt agctggagaa gataatgtac 5160 5220 tcaccaactc cactctatgc aacccagtct getetgeeca tgecagtcag acgtgaatet 5280 tacacctgga ttcagattga tgaatctaca acatcaccca ctccatgctt ccttctaaat cagcagttct agcctgaatg acagatgcta cccaagtctc atctagttag ccctgtccgg 5340 agtaaccctg accttgagga ttagaccagg atgcacatcc tgcaccagtt ccctttgtcc 5400 acctgacttc atcccacccg ggccatagcc catgctcagg ctccaccctc catgcacaaa 5460 gctggctttt ccagcttcct tcacctgtat cagacacaaa tagcaaaagg ggtccacgtg 5520 5580 cctaggtccc atcacaagac catgtgcggt agtttggaaa acagtctcca cttgaggctc agatagnttg gaatcttggc tctcatgtag ttgtactgat tagatcagtt taggaagtat gacctttntg gaaagaacat ataactggga ggggctttga gatgtaaagg ccccacacaa 5700 ttcctagttc aaactntact gcctgctcaa agcttgaggc atgaactctc actgttcctg 5760 5820 atgtcatggt tectgtetge ttecacaatt cectateatt aggggteeet tteetteetg 5880 gaattttaag tataaataaa cncttctttn taaaacaaca agaacaacaa atctgacnct gataatggat tttaaggcgt cttctctgga taagaaaaaa aaaagaatat atttgcatag 5940 gtgctgtatt acttttgtca ttggtataac ctgactggaa gcaacttaaa ggaagaagaa 6000 tgtatcttga tttgtagatt aagagcacca tgactaagaa ggcatagcag cacaggtgca 6060 ccagcaagaa cataggctgc tagctcagat ctctgtagat atgggaacag ggcaggaagc 6120 tagtagteta taaaceteag gacceatece atggagttee ttgtetteea gtgatgteet 6180 gtgtcttaaa gtttcacagt tcccacagca gcacctgccg tctgggaacc aacctgtggt 6240 6300 ggatatttta caacgtgata ggcatatttt gtctctagcc ctgtaggttt atagccatcc 6360 tatacttcag tttatctagt ccacctcagt ctgatggtct tatagttcca acacttcaaa actacaaagt cttaagggcc atgggctcgg gtttattaga gcagtaacac ctctactagc 6420 tttctgtgtt acccactcct cttaaggtct ggttgaaatc ctaataggaa gcagcttgag 6480 aggagggttt attgtggccc atactttgtt ggtacattct atcatgcaag ggtggcactg 6540 tgatacagcc gaggccatcc gaggatggta ctgttggctt acatctgggt gggacaggaa 6600 atggtaattc tcaaggccca cctgcttggt gacttctttc agttaagccc catactctaa 6660 atcctctaca acctcccaac ataatgccac cagctgggga tcagctgttg acagtgctgg 6720 cccaggggag cagtttaaat ccagaccagg ggacctgaaa acagagaact gcagaggggc 6780 6840 tgtgggactt tataccagct ttgcagacaa atcacggcat ttctttgtga gcttggttca taaacaaata tatattctcc tataggctcc tttagtgggt gtttcatatc cacaaatttg 6900 ttcagaaaaa cactgtgttt tatgctagct gtgtaggaga taataccgct gggagtcact 6960 7020 tgagcatgga taagtgacat agttcgtcct catgagtccc tgtcctgttt ctgtattatg 7080 tttacttgat gagtttagtt tgtcagttgg ccaccaatta aaaagtatca ttttattttt tttacaatac tcagttctca agttaggagt tttgttatta tatggcttca atattcacat 7140 tttaaccttt ccaggagtta agtataaaaa cttatatcaa ctgttgactt agtaaatatc 7200 7260 tattacagat actatattct tcttagttta tatcatgaat atgaggttgc ttaaagtaag 7320 tgatgtaaaa tacactaggg gatgcttata aaatggaatg ttgtgagttt tttgaaacac gagtactaaa ttcataagtt tttaaatagt tacactgtta gcttcagtac tgctagatac 7380 atgtctataa tggctgaaga gtggagcttg gatattataa gtgtactctg tatattcatg 7440 7500 cagacatata gcagattcca ctagtatgtg tggttaatat gtgctaataa aaatttaata 7560 caaaagtcat gttttattac tgggaaccag aggggttggt tgtgctgatt ttaagtcagt gactattagc atattctaag aaacagtttt naggatttta aagattggct ttaccataaa 7620 tgtagagcta tgttttacta taatccatat tatggtcggc cttaattcaa tctctgcagt 7680 7740 ttggttactc tgctcaaagt gaaggtcatt tataaatgat acacattttc tcaccatagg 7800 aaatactacc tggccaataa cagagttaga attgctaaat tgatggtacc aacaatggac 7860 tcaacacaaa ctaaagttta tttatgccca cagatgtatc ttgtgatttt cccagaggga acaaggtata atgcaacata cacaaaactc ctttcagcca gtcaggcatt tgctgctcag 7920 cggggtaagt aaagatttaa ctgtattcag aaaaacactt ttttaagaag agtgatcttt 7980 8040 gtttccttca gagtcatact aaagaatatg cgtttcttgt aagagctaag tgagagaata tccgatcttc tacagagtta ggtatattct tattagtctg tgtctgagag gttagagacg 8100 8160 caggettget atggeacatt teccatgetg tgaattgagt taaaaatgta ggtaaatgat atccccaaga aagtatactt ttnggagtga ctcagtataa agcctggtgt tataacataa 8220 acacgcacgt gcggatgtat atgtagcaca tatgtaaaca caggtatatg canttgtaat 8280 aagaaagtgg aggtcggggc ccactgcagg caagtctttt agtgatgctg agctaatgct 8340 gagaggtaga aagaccaaga aggctggagt tgctcattcg gcaaaggtca gagctcactg 8400 tgtgccataa ctcgagtgtt ctgtctccct tttgatacag ttttcttgtt tttaattatt 8460 agtttttaca attatcccat aaaatgtggg ctcattgtgg tcatcgtttt cataaagtcc 8520 ttcaagtata cacccagcaa gtatctaaat acactgggaa gaatcagtca gctgatggct 8580 8640 tgaagtttca ggacatctag tgccacatca tgcttcagaa ccgacctgca cttagtcagg 8700 gtcatattca tgccacgtga agacgagagg aggccatgcc gtctgactta ggatggaaat

8760 ttccttcgag caaacacgaa cgggctaggt cttagttata ggcatagtgt ctgtggttat actaggcaga cattagtgga ctgggtgtta gaaggtacag acaggcaaga atttgctgta 8820 8880 gatttgtttc cctcatgtgt tgacaccaca tctaacctgc tttttgagct tctagtccta 8940 ataatctcat aaaaatactg gttgaaccag aaatggtgtt gcaaagctat gatcccagct 9000 cctgggatct agggtgggag gatcataaat ttgaggccag cttgggtctg tcttagagaa aaaagaaaaa taaaaagtct ggtcaaggta acatggagcc tggaagtttc acagggtgat 9060 tctgtaaagg tcctgagaca agatggcctc tagtggcgaa tgacttagct gacaangaaa 9120 actttcccag cttggttgac ttttcagact tcatacaagt ttgtgaataa attacactcc 9180 ttctgccctt gggactgaac tcagatatgt ggttgtggga atggctttct ttcccacacc 9240 9300 accetgeatt ttaaaaatte ttetgtagae agteecacea teetgtaget gttetteett atgtcgccac tttccctgga gagaggcagt gcagacttca acccgcttct ccctagtcgc 9360 tgttcatagc acatcgaaag acctagtgct tcctgtgaaa ttgtaagtac atcctggagt 9420 9480 ccaggagagg aggaagccga acagagtgga gggaatgctg agttctgtcc taagaaagac tgcgtgctta gcaagatgct gctgctctcc tgtcgtgtct ttcttgtcag aacttatcaa 9540 agagaagget egeagtgggt cataatette ecaaggacea geetteecag ettetegeag 9600 9660 catatctcat tcatgtagat gtttaatgga tatgtgtcaa tggggttgac ctaagtgaga tggcaatgta tgtgagcatt ctaggtgtga ggttatggca ttaaacttta atttccgtct 9720 atttgtggta gttgataagt aatttagatg ttgactttca tgtattccta attatgacca 9780 cattgaatct acctgctttc taggccttgc agtattaaaa cacgtactga caccaagaat 9840 aaaggccact cacgttgctt ttgattctat gaagagtcat ttagatgcaa tttatgatgt 9900 9960 cacagtggtt tatgaaggga atgagaaagg ttcaggaaaa tactcaaatc caccatccat gactggtaag tccgtatttc catagaagct gaatagtaca tggtacaggt aagataaact 10020 cttgtttgtt cgctttgctt agcttggttc agtttggttt tcagtagagg gttccactat 10080 gaagetetgg etggeeggga acteactatg tagaceaage tggeettggg etceactaea 10140 cccagcacca atcacccact cttatncttt tatgentttt tgtttttgct ttgagettte 10200 tttataacat gtttgggaag gacattgtca ttatttacaa gaagaaatat ggtcttttcc caacatgcta gaatttaaag actcagaact cttgcctttg tcagtgacaa agtgagaatg 10320 gctgtgaagt gacgtggctt tgagtgagaa tagttcaggt aactatagcc acagactcaa 10380 catttgaaca tgggaacagg tgagaacgga gtgatggaag attctggccc ctttcagaga 10440 10500 gctgcagact aaagagatct cttataatcg cagtactaag gaggaagaag cagaagatga tgactacagg gccaggctga acaatctagt aaaatcctaa gtcaggaagt cagggctgag 10620 gtgcagctca gtaggagagt tgttgtctgc cctacacaag gcctggattt agctcccagt 10680 agcaacgaag ggaggcgagg gtgggcaaaa tcgaacactt actcttggag actcccttta tgaatattac cacactccag taaatactct ccagagattt cagatgagat tctgcttcct 10800 ggtaaacagg aggccaagaa tattatgtca cactgaacat gggatggaag acatgttctg 10860 aggaatgtct gcactccagt gtgatgaaga cttgaagttt agggacattt tccctccctg 10920 gccccactca ccccatctgt attgagtatt cccctagtgc tcatctttat ttgtatgtta 10980 actttcagga aggggaagca gattgatatt caaacccagc cagttttctt aaatactttg 11040 tggatgggat tggctttgac agtaaatgag gaaatgtaaa atgtaaaaga ttctaatttt 11100 taatatttta aaggtgaggt tttctgttag tacgcagagt gagaggtttc ttactgatgt 11160 ctgcgtacct agaggaagga tggctacttc tccaaggctt gctgttagaa gtcagtgaca 11220 tgggcttaac aagagatatg tgctaatgag gttttaattt cagcttaata ctgcaaatca 11280 taagtgcata gctttattgt tttaaattct tttagtctta atgtttcatt tttaccataa 11340 gttactttgt ataatcacaa attctaaact agtaagacgt gaaattttct tcttctttgt 11400 tagagtttct ctgcaaacag tgcccaaaac ttcatattca ctttgatcgt atagacagaa 11460 atgaagttcc agaggaacaa gaacacatga aaaagtggct tcatgagcgc tttgagataa 11520 aagataggta agtggtaaga gctccagcat ttagaaagtg cagttcaacc aaattttact 11580 ctcagatcct gcttgaaagg agtcttttta tcttcattat ttagtaaata ctaatcatac 11640 11700 ctgcatagac aagaccacat atacttaaat gtagcatgtt tcatggtgcg ttacccttgt 11760 ttaacaatta agtttaacat cctacatcag tttgcctgtt gatttctgta ccatgacaac tcaacacagc gatgcgttta ttccaaagtc gatagcacag caaaagtgaa actaaagtct 11820 gtattgtttc aagaatgctt tttgtgaact cgggttaaat cttattctat cctttcgtgt 11880 tcacattgta cattttcatg agtcactata qaaatcatga catggtggcc tacctgcagt 11940 gtttgctgga cagtaggctg ctgtgtgata agagcctttc ctcttcagct acacggggga 12000 cacgaggett tggggttcaa gactgaagca cgggtgagca caacacettt gtgttgtggg 12060 aaggaaggga attgttcttt tcataatgaa attgtcccct ttcttgagtt agtagaaagt 12120 attacaagga tagagagttg aaatgaagct ttatattaga tttatgcctt gtgttgtcac 12180 gtgtttctac ctgacataac ttttcaaccc agccgctcag gattattttg atgatgggaa 12240 caatgtaaga aggcctatgt atcggtaact cactgttgta gctctgtgga ancggntcnn 12300 caggcagtag ggacgcttct gtgcttttgt gcctgtattg ctgttagaat cttacagagg 12360

aggatgaatg aatgaccett tttatttete ttgtetgett ttetaatttt atgggaataa 12420 gaacttttgg taggtctctg tcactggcct cttgttgtga agagacacct tgagcaaagc 12480 aactcttctg agagaaagca tttagttggg gaattcctta cagcttcaga ggttgagttc 12540 gttttcatca tgctgaggac agggaggcac tcaggcagga gaagtagttg agagccacat 12600 tetgatetae aggeagagag agacagaetg ageetggeat gggttettgg aaceteaaag 12660 cetetcatee etaceceate tecegaceee tatacacaet tectecaaca aggetacace 12720 ttctaatcct tcttaaagag tcaccacatc cagcgactaa gcattcagat atgtgaacct 12780 gtcggagcct ttcttactca gatcacctca ggaggaaaac tcctatgcta taagaatttc 12840 ttttctttcg catctttgaa agcttgtttt tgtgtgatta gatcctggcc tcacacatgc 12900 teggeaatea ttttactgtt gageteeage eteageegtt tteattgget tatgggatge gagccatggg agagaagcta gaaggccttt cgttttatga gtcgggttgg tggaaccact 13020 tacagatgga agatttacaa acaaaaatga agctggggcc atcaaggctc agcactcgct 13080 getettecag agagtteagg tteattete agtaaceaea tggtggettt gtaaatgtaa 13140 cttcatattc aatgaccctg acaccctctt ctggcctctg tgggcaccag acacaatcat 13200 ggtatacaga cacacact agccaacacc catctacata aaagtatata anacatatct 13260 ttatcttaaa aatccccgaa gtcctcatta aatatcttag atccccgccg tgttttgatt 13320 tttgtttccc acgtggtgag gatataatat catgtccaaa ctgtaaggag tgaatgccct 13380 cccgtgcctc tcggacacct ctgcactcat ccaagttttc taaggagctg tacttgctca 13440 gcaagtactc aatacctaat aaatggttta tgtttgtttc aacaccaaaa atgtccaaaa 13500 ctgaaagatc aattctgttg ttttccttct ggccataggt tgctcataga gttctatgat 13560 tcaccagatc cagaaagaag aaacaaattt cctgggaaaa gtgttcattc cagactaagt 13620 gtgaagaaga ctttaccttc agtgttgatc ttggggagtt tgactgcggt catgctgatg 13680 acggagtecg gaaggaaact gtacatgggc acctggttgt atggaaccet cettggetge 13740 ctgtggtttg ttattaaagc ataagcaagt agcaggctgc agtcacagtc tcttattgat 13800 ggctacacat tgtatcacat tgtttcctga attaaataag gagttttctt gttgttgttt 13860 ttttgtttt gttttgttct gttttaagcc ttgatgattg aacactggat aaagtagagt 13920 ttgtgaccac agccaacatg catttgattt ggggcaaaca catgtggctt ttcaggtgct 13980 ggggttgctg gagacatgga agctaagtgg agtttatgct gnttttttt ttttttnaa 14040 tgttttcatg aattaatgtc cacttgtaaa gattattgga tactttctgt aattcagaag 14100 gttgtatttt aacactagtt tgcagtatgt ttcgctatat tggttatctt ccatttgact 14160 acttggcagc tcagactctt aatactaaag tattttacat tttgaagcta tgtgatactg 14220 gttttttgtt gttgttgttg ttgttaattt ctgaaagtca atgaaagaca ctgtaatgat 14280 gcgttaagat gttccaagaa aaaggtgaga attattcatg gcaaaaaaga tctgtctagt 14340 gtatatttt attatattgc tctatttagc taattttctt tatatttgca aaataatgaa 14400 catttttaat atttattaaa atgcttgatt tgcatacccc cgattctaca gagaataatg 14460 tgtaaagtgt cagaatagac ttgaagctct gctgtgactc agtctccttt gtcagagctt 14520 ctagtagccc agetactgag ctgctttgtt agtacctcca gcacctgagc cgttaagtac 14580 ttataaatgc aagggacccg ttatcttcat atcggaatag acatgaacag agctctaagg 14640 cgatgaaagt ctgccagcat cctctctgtc ctcgcacgtg ccttctgcct ggctccattt 14700 getttggcac tgcgttcgat ctagagtgta ggtgctcact gettatttca geeetggete 14760 tg:ggttttg tgtcctccag tggtgctgtt cactgttggg gtgcaggtgg tgctgccctg 14820 actcagaggg gcageteect ggeteetgag ggtgageett ettggetaet acagaagtat 14880 tgtgcgtttg tgtatggcaa gaaccatcag gattggataa atgtgttatt tctctttgat 14940 ttccatggag ccacactgtt ggtacatgtc ccctgtgaac agagctacct ttcaggagca 15000 catcatactg togtgagtca oggcacggtg tgtootgtga gaagaggott totaacgtgt 15060 gatttgccgt gtttctatgt tgtgatttaa gcgtgattgc ctactagtca ttcaaggtaa 15120 catttetgea aattteatae agatttttgt cacaaaatta etataceaat gatetagttg 15180 aaatagacca attgaatcac aataaataat tttttttaat tgagggaaaa tttgcttctt 15240 gttttttcaa agccagaaaa cgagccattt caaacatctt tgaagagtca tgtgctgtca 15300 cttgttttct atgtgttagt gtctatattc atgtatggat acacatgaac atgtatattc 15360 atacacaca gccaatagaa tataacagcc taaaaacaat ccagcttgtg tatcatgtta 15420 ctgtgctgaa ttgtaatggt ttttacttac aaagtgaggc taaaatcgat ttcatgtctt 15480 tgttaaatac gttttttca gcaatcctat tagagcttat tttgaccaga tcaaaataag 15540 tacaagttca gagactttaa atatggctga ggtctagagc gatagctcag tagttaggaa 15600 cacatgccac tetttcaagg gettcagtte ecagcactca tatggagget cacagaagge 15660 tggaattcca getteatgga attggacaca teetetaget teeatggate tgtetgtetg 15720 tetetecett etetetet etetetet etetetet etetet 15766 <210> 74 <211> 354

<212> PRT

<213> Mus musculus

22

```
<400> 74
Met Arg Tyr Leu Leu Pro Ser Val Leu Leu Gly Ser Ala Pro Thr
       5
                            10
Tyr Leu Leu Ala Trp Thr Leu Trp Arg Val Leu Ser Ala Leu Met Pro
                            25
                                    3.0
Ala Arg Leu Tyr Gln Arg Val Asp Asp Arg Leu Tyr Cys Val Tyr Gln
                       40
                                         45
      35
Asn Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
                    55
                                    60
Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Val Ile Tyr Leu Ala
              70
                                   75
Asn His Gln Ser Thr Val Asp Trp Ile Val Ala Asp Met Leu Ala Ala
                               90
                                                 95
             85
Arg Gln Asp Ala Leu Gly His Val Arg Tyr Val Leu Lys Asp Lys Leu
          100
                           105
Lys Trp Leu Pro Leu Tyr Gly Phe Tyr Phe Ala Gln His Gly Gly Ile
                       120
                                         125
       115
Tyr Val Lys Arg Ser Ala Lys Phe Asn Asp Lys Glu Met Arg Ser Lys
                   135
                               140
Leu Gln Ser Tyr Val Asn Ala Gly Thr Pro Met Tyr Leu Val Ile Phe
        150
                                155
Pro Glu Gly Thr Arg Tyr Asn Ala Thr Tyr Thr Lys Leu Leu Ser Ala
                             170
             165
                                        175
Ser Gln Ala Phe Ala Ala Gln Arg Gly Leu Ala Val Leu Lys His Val
                           185
                                           190
          180
Leu Thr Pro Arg Ile Lys Ala Thr His Val Ala Phe Asp Ser Met Lys
                        200
                                          205
Ser His Leu Asp Ala Ile Tyr Asp Val Thr Val Val Tyr Glu Gly Asn
                    215
                                      220
Glu Lys Gly Ser Gly Lys Tyr Ser Asn Pro Pro Ser Met Thr Glu Phe
                  230
                          235
Leu Cys Lys Gln Cys Pro Lys Leu His Ile His Phe Asp Arg Ile Asp
                               250
           245
Arg Asn Glu Val Pro Glu Glu Gln Glu His Met Lys Lys Trp Leu His
                 265 270
          260
Glu Arg Phe Glu Ile Lys Asp Arg Leu Leu Ile Glu Phe Tyr Asp Ser
       275 280
                                          285
Pro Asp Pro Glu Arg Arg Asn Lys Phe Pro Gly Lys Ser Val His Ser
                     295
                                       300
Arg Leu Ser Val Lys Lys Thr Leu Pro Ser Val Leu Ile Leu Gly Ser
                310 315
305
Leu Thr Ala Val Met Leu Met Thr Glu Ser Gly Arg Lys Leu Tyr Met
             325
                              330
                                              335
Gly Thr Trp Leu Tyr Gly Thr Leu Leu Gly Cys Leu Trp Phe Val Ile
                  345
Lys Ala
<210> 75
<211> 22
<212> DNA
<213> Mus Musculus
<220>
<221> misc_binding
<222> 1..22
<223> amplification oligonucleotide g34292.pu
<400> 75
attaaaacac gtactgacac ca
<210> 76
<211> 19
 <212> DNA
```

<213> Mus Musculus

<220>

WO 99/32644 PCT/IB98/02133 48

<221> misc_binding	
<222> 119	
<223> amplification oligonucleotide g34292.rp	
<400> 76	1.0
agtcatggat ggtggattt	19
<210> 77	
<21.1> 26	
<2.2> DNA	
<213> Homo Sapiens	
<220>	
<221> misc_binding	
<222> 126	
<223> amplification oligonucleotide BOXIed	
<400> 77	
aatcatcaaa gcacagttga ctggat	26
<210> 78	
<211> 33	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> misc_binding	
<222> 133	
<223> amplification oligonucleotide BOXIIIer	
<400> 78	
ataaaccacc gtaacatcat aaattgcatc taa	33
<210> 79	
<211> 22	
<212> DNA	
<213> Mus Musculus	
<220>	
<221> misc_binding	
<222> 122	
<223> sequencing oligonucleotide moPGrace3S473	
<400> 79	
gagataaaag ataggttgct ca	22
<210> 80	
<211> 19	
<212> DNA	
<213> Mus Musculus	
<pre><220> <221> misc_binding</pre>	
<222> misc_binding <222> 119	
<pre><222> 119 <223> sequencing oligonucleotide moPGrace3S526</pre>	
<400> 80	
	19
aagaaacaaa tttcctggg	
<210> 81	
<211> 18	
<212> DNA <213> Mus Musculus	
<2213 Mus Musculus <220>	
<221> misc_binding	
<222> 118	
<223> sequencing oligonucleotide moPGrace3S597	
<400> 81	18
tcttggggag tttgactg	10
<210> 82	
<211> 18	
<212> DNA	
<213> Mus Musculus	
<220>	
<221> misc_binding	

<222> <223>	118 sequencing oligonucleotide moPGrace5R323	
<400>		
	cggtg tagttctc	18
<211>		
<212>		
	Mus Musculus	
<220>		
	misc_binding	
	117	
-2222	sequencing oligonucleotide moPGrace5R372	
<223>	sequencing offgondereotide morefules.	
<400>		17
	aagec ggtcgtc	
<210>		
<211>	17	
<212>		
<213>	Mus Musculus	
<220>		
<221>	· misc_binding	
<222>	. 117	
<223>	sequencing oligonucleotide moPGrace5R444	
<400>		
	cagca ggtaggt	17
<210>		
<211>		
<212>	· · · · · · · · · · · · · · · · · · ·	
	Mus Musculus	
<220>		
	<pre>> misc_binding</pre>	
<222>	> 119 > sequencing oligonucleotide moPGrace5R492	
<400>		19
	ggtagc gcatagagt	
<210>		
<211>		
	> DNA	
	> Mus Musculus	
<220>		
	> misc_binding	
<222>	> 127	
	> amplification oligonucleotide moPG13LR2	
<400>		27
	acaatg tgatacaatg tgtagcc	
<210>		
<211:		
<212	> DNA	
	> Mus Musculus	
<220:		
	> misc_binding	
<222	> 118	
<223	> amplification oligonucleotide moPG15	
<400	9> 87	18
tggc	gagccg agaggatg ,	19
<210)> 88	
<211	L> 36	
<212	2> DNA	
	3> Mus Musculus	
<220		
	l> misc_binding	
	2> 136	

•	oligonucleotide morgiobami	
<400> 88		36
cgtggatccg gaaacaatg	t gatacaatgt gtagcc	30
<210> 89		
<211> 27		
<212> DNA		
<213> Mus Musculus		
<220>		
<221> misc_binding		
<222> 127		
<223> amplification	oligonucleotide moPG15Eco1	
<400> 89		
cgtgaattct ggcgagccg	a gaggatg	27
<210> 90	• • • •	
<211> 20		
<212> DNA		
<213> Mus Musculus		
<220>		
<221> misc_binding		
<222> misc_binding <222> 120		
2222 IZu	oligonucleotide moPG1RACE3.18	
	Oligonacieotiae moi olialozo la	
<400> 90	•	20
ctgccagaca ggatgccct	,a	
<210> 91		
<211> 23		
<212> DNA		
<213> Mus Musculus		
<220>		
<221> misc_binding		
<222> 123	1 1 2 1 1 1 2 m = PO1 PAGE 2 6 2	
	oligonucleotide moPG1RACE3.63	
<400> 91		23
acaagttaaa atggcttco	eg ctg	23
<210> 92		
<211> 18	•	
<212> DNA		
<213> Mus Musculus		
<220>		
<221> misc_binding		
<222> 118		
<223> sequencing ol:	igonucleotide moPG1RACE3R94	
<400> 92		
caaatgcatg ttggctgt		18
<210> 93		
<211> 20		
<212> DNA		
<213> Mus Musculus		
<220>		
<221> misc_binding		
<222> 1 20		
<223> amplification	oligonucleotide moPG1RACE5.276	
<400> 93	-	
gcaaatgcct gactggct	.ga	20
<210> 94	,	
<210> 34 <211> 22	·	
<211> 22 <212> DNA		
<213> Mus Musculus		
<220>		
<221> misc_binding		
<221> misc_binding <222> 122		
<223> amplification	n pligonucleotide moPG1RACE5.350	

<400> 94	22
aatcaaaagc aacgtgagtg gc	22
<210> 95	
<211> 20	
<212> DNA	
<213> Mus Musculus	
<220>	
<221> misc_binding	
<222> 120	
<223> amplification oligonucleotide moPG3RACE2	
<400> 95	20
tqqqcacctg gttgtatgga	20
<210> 96	
<211> 20	
<212> DNA	
<213> Mus Musculus	
<220>	
<221> misc_binding	
<222> 120	
<223> amplification oligonucleotide moPG3RACE2n	
<400> 96	
teettagetg cetgtggttt	20
<210> 97	
<211> 21	
<212> DNA	
<213> Mus Musculus	
<220>	
<221> misc_binding	
<222> 121	
<223> sequencing oligonucleotide moPG3RACES20	
<400> 97	0.1
gatggctaca cattgtatca c	21
<210> 98	
<211> 24	
<212> DNA	
<213> Mus Musculus	
<220>	
<221> misc_binding	
<222> 124	
<223> sequencing oligonucleotide moPG3RACES5	
<400> 98	2.4
tcctgaatta aataaggagt tttc	24
<210> 99	
<211> 24	
<212> DNA	
<213> Mus Musculus	
<220>	
<221> misc_binding	
<222> 124	
<223> sequencing oligonucleotide moPG3RACES90	
<400> 99	2.4
gtttgttatt aaagcataag caag	24
<210> 100	
<211> 216	
<212> DNA	
<213> Homo sapiens	
<400> 100	60
ctgctgtccc tggtgctcca cacgtactcc atgcgctacc tgctgcccag cgtcgtgctc	120
ctgggcacgg cgcccaccta cgtgttggcc tggggggtct ggcggctgct ctccgccttc	180
ctgcccgccc gcttctacca agcgctggac gaccggctgt actgcgtcta ccagagcatg	216
gtgctcttct tcttcgagaa ttacaccggg gtccag	∠⊥0

```
<210> 101
<211> 70
<212> DNA
<213> Homo sapiens
<400> 101
                                                                        60
atattgctat atggagattt gccaaaaaat aaagaaaata taatatattt agcaaatcat
                                                                        70
caaaggacag
<210> 102
<211> 116
<212> DNA
<213> Homo sapiens
<400> 102
ttgactggat tgttgctgac atcttggcca tcaggcagaa tgcgctagga catgtgcgct
                                                                        60
                                                                       116
acgtgctgaa agaagggtta aaatggctgc cattgtatgg gtgttacttt gctcag
<210> 103
<211> 90
<212> DNA
<213> Homo sapiens
<400> 103
catggaggaa tctatgtaaa gcgcagtgcc aaatttaacg agaaagagat gcgaaacaag
                                                                        60
                                                                        90
ttgcagagct acgtggacgc aggaactcca
<210> 104
<211> 91
<212> DNA
<213> Homo sapiens
<400> 104
atgtatcttg tgatttttcc agaaggtaca aggtataatc cagagcaaac aaaagtcctt
                                                                        60
                                                                        91
tcagctagtc aggcatttgc tgcccaacgt g
<210> 105
<211> 159
<212> DNA
<213> Homo sapiens
<400> 105
gccttgcagt attaaaacat gtgctaacac cacgaataaa ggcaactcac gttgcttttg
                                                                        60
attgcatgaa gaattattta gatgcaattt atgatgttac ggtggtttat gaagggaaag
                                                                       120
                                                                       159
acgatggagg gcagcgaaga gagtcaccga ccatgacgg
<210> 106
 <211> 124
 <212> DNA
 <213> Homo sapiens
 <400> 106
 aatttctctg caaagaatgt ccaaaaattc atattcacat tgatcgtatc gacaaaaaag
                                                                        60
 atgtcccaga agaacaagaa catatgagaa gatggctgca tgaacgtttc gaaatcaaag
                                                                        120
                                                                        124
 ataa
 <210> 107
 <211> 4342
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> polyA_signal <222> 325..330
 <223> AATAAA potential
 <221> polyA_signal
 <222> 694..699
 <223> AATAAA potential
 <221> polyA_signal
 <222> 828..833
 <223> AATAAA potential
 <221> polyA_signal
 <222> 1821..1826
 <223> AATAAA potential
```

```
<221> polyA_signal
<222> 2480..2485
<223> AATAAA : potential
<221> polyA_signal
<222> 2800..2805
<223> AATAAA potential
<221> polyA_signal
<222> 4264..4269
<223> AATAAA potential
<221> polyA_signal <222> 4320..4315
<223> AATAAA
<400> 107
gatgcttata gaattttatg agtcaccaga tccagaaaga agaaaaagat ttcctgggaa
                                                                      120
aagtgttaat tocaaattaa gtatoaagaa gaotttacca toaatgttga tottaagtgg
tttgactgca ggcatgctta tgaccgatgc tggaaggaag ctgtatgtga acacctggat
                                                                      180
atatggaacc ctacttggct gcctgtgggt tactattaaa gcatagacaa gtagctgtct
                                                                      240
ccagacagtg ggatgtgcta cattgtctat ttttggcggc tgcacatgac atcaaattgt
                                                                      300
                                                                      360
ttcctgaatt tattaaggag tgtaaataaa gccttgttga ttgaagattg gataatagaa
                                                                      420
tttgtgacga aagctgatat gcaatggtct tgggcaaaca tacctggttg tacaacttta
gcatcggggc tgctggaagg gtaaaagcta aatggagttt ctcctgctct gtccatttcc
                                                                      480
tatgaactaa tgacaacttg agaaggctgg gaggattgtg tattttgcaa gtcagatggc
                                                                      540
tgcatttttg agcattaatt tgcagcgtat ttcacttttt ctgttatttt caatttatta
                                                                      600
caacttgaca gctccaagct cttattacta aagtatttag tatcttgcag ctagttaata
                                                                      660
tttcatcttt tgcttatttc tacaagtcag tgaaataaat tgtatttagg aagtgtcagg
                                                                      720
atgttcaaag gaaagggtaa aaagtgttca tggggaaaaa gctctgttta gcacatgatt
                                                                      780
ttattgtatt gcgttattag ctgattttac tcattttata tttgcaaaat aaatttctaa
                                                                      840
tatttattga aattgcttaa tttgcacacc ctgtacacac agaaaatggt ataaaatatg
                                                                      900
agaacgaagt ttaaaattgt gactctgatt cattatagca gaactttaaa tttcccagct
                                                                      960
ttttgaagat ttaagctacg ctattagtac ttccctttgt ctgtgccata agtgcttgaa
                                                                     1020
aacgttaagg ttttctgttt tgttttgttt ttttaatatc aaaagagtcg gtgtgaacct
                                                                     1080
tggttggacc ccaagttcac aagattttta aggtgatgag agcctgcaga cattctgcct
                                                                     1140
agatttacta gcgtgtgcct tttgcctgct tctctttgat ttcacagaat attcattcag
                                                                     1200
aagtegegtt tetgtagtgt ggtggattee caetgggete tggteettee ettggateee
                                                                     1260
gtcagtggtg ctgctcagcg gcttgcacgt agacttgcta ggaagaaatg cagagccagc
                                                                     1320
ctgtgctgcc cactttcaga gttgaactct ttaagccctt gtgagtgggc ttcaccagct
                                                                     1380
actgcagagg cattttgcat ttgtctgtgt caagaagttc accttctcaa gccagtgaaa
                                                                     1440
 tacagactta attcgtcatg actgaacgaa tttgtttatt tcccattagg tttagtggag
                                                                     1500
 ctacacatta atatgtatcg ccttagagca agagctgtgt tccaggaacc agatcacgat
                                                                     1560
 ttttagccat ggaacaatat atcccatggg agaagacctt tcagtgtgaa ctgttctatt
                                                                     1620
tttgtgttat aatttaaact tcgatttcct catagtcctt taagttgaca tttctgctta
                                                                     1680
 ctgctactgg atttttgctg cagaaatata tcagtggccc acattaaaca taccagttgg
                                                                     1740
 atcatgataa gcaaaatgaa agaaataatg attaagggaa aattaagtga ctgtgttaca
                                                                     1800
 ctgcttctcc catgccagag aataaactct ttcaagcatc atctttgaag agtcgtgtgg
                                                                     1860
 tgtgaattgg tttgtgtaca ttagaatgta tgcacacatc catggacact caggatatag
 ttggcctaat aatcggggca tgggtaaaac ttatgaaaat ttcctcatgc tgaattgtaa
                                                                      1980
                                                                      2040
 ttttctctta cctgtaaagt aaaatttaga tcaattccat gtctttgtta agtacaggga
                                                                      2100
 tttaatatat tttgaatata atgggtatgt tctaaatttg aactttgaga ggcaatactg
 ttggaattat gtggattcta actcatttta acaaggtagc ctgacctgca taagatcact
                                                                      2160
 tgaatgttag gtttcataga actatactaa tcttctcaca aaaggtctat aaaatacagt
                                                                      2220
 cgttgaaaaa aattttgtat caaaatgttt ggaaaattag aagcttctcc ttaacctgta
                                                                      2280
 ttgatactga cttgaattat tttctaaaat taagagccgt atacctacct gtaagtcttt
 tcacatatca tttaaacttt tgtttgtatt attactgatt tacagcttag ttattaattt
                                                                      2400
 ttctttataa gaatgccgtc gatgtgcatg cttttatgtt tttcagaaaa gggtgtgttt
                                                                      2460
 ggatgaaagt aaaaaaaaaa ataaaatctt tcactgtctc taatggctgt gctgtttaac
                                                                      2520
 attttttgac cctaaaattc accaacagtc tcccagtaca taaaataggc ttaatgactg
                                                                      2580
 gccctgcatt cttcacaata tttttcccta agctttgagc aaagttttaa aaaaatacac
                                                                      2640
                                                                      2700
 taaaataatc aaaactgtta agcagtatat tagtttggtt atataaattc atctgcaatt
 tataagatgc atggccgatg ttaatttgct tggcaattct gtaatcatta agtgatctca
                                                                      2760
 gtgaaacatg tcaaatgcct taaattaact aagttggtga ataaaagtgc cgatctggct
                                                                      2820
 aactettaca ccatacatac tgatagtttt tcatatgttt catttccatg tgatttttaa
                                                                      2880
```

```
aatttagagt ggcaacaatt ttgcttaata tgggttacat aagctttatt ttttcctttg
                                                                    2940
ttcataatta tattctttga ataggtctgt gtcaatcaag tgatctaact agactgatca
                                                                    3000
tagatagaag gaaataaggc caagttcaag accagcctgg gcaacatatc gagaacctgt
                                                                    3060
ctacaaaaaa attaaaaaaa attagccagg catggtggcg tacactgagt agtttgtccc
                                                                    3120
agctactcgg gagggtgagg tgggaggatc gcttcagccc aggaggttga gattgcagtg
                                                                    3180
agccatggac ataccactgc actacagcct aggtaacagc acgagacccc aactcttaga
aaatgaaaag gaaatataga aatataaaat ttgcttatta tagacacaca gtaactccca
                                                                    3300
                                                                    3360
gatatgtacc acaaaaaatg tgaaaagaga gagaaatgtc taccaaagca gtattttgtg
                                                                     3420
tgtataattg caagcgcata gtaaaataat tttaacctta atttgttttt agtagtgttt
agattgaaga ttgagtgaaa tattttcttg gcagatattc cgtatctggt ggaaagctac
                                                                     3480
aatgcaatgt cgttgtagtt ttgcatggct tgctttataa acaagatttt ttctccctcc
                                                                     3540
ttttgggcca gttttcatta cgagtaactc acactttttg attaaagaac ttgaaattac
                                                                    3600
gttatcactt agtataattg acattatata gagactatgt aacatgcaat cattagaatc
aaaattagta ctttggtcaa aatatttaca acattcacat acttgtcaaa tattcatgta
                                                                     3720
attaactgaa tttaaaacct tcaactatta tgaagtgctc gtctgtacaa tcgctaattt
                                                                     3780
actcagttta gagtagctac aactcttcga tactatcatc aatatttgac atcttttcca
                                                                     3840
                                                                     3900
atttgtgtat gaaaagtaaa tctattcctg tagcaactgg ggagtcatat atgaggtcaa
agacatatac cttgttatta taatatgtat actataataa tagctggtta tcctgagcag
gggaaaaggt tatttttagg aaaaccactt caaatagaaa gctgaagtac ttctaatata
                                                                     4020
                                                                     4080
ctgagggaag tataatatgt ggaacaaact ctcaacaaaa tgtttattga tgttgatgaa
acagatcagt ttttccatcc ggattattat tggttcatga ttttatatgt gaatatgtaa
                                                                     4140
gatatgttct gcaattttat aaatgttcat gtcttttttt aaaaaaggtg ctattgaaat
totgtgtoto cagcaggcaa gaatacttga ctaactcttt ttgtctcttt atggtatttt
                                                                     4260
cagaataaag tctgacttgt gtttttgaga ttattggtgc ctcattaatt cagcaataaa
                                                                    4320
                                                                     4342
ggaaaatatg catctcaaaa at
<210> 108
<211> 62
<212> DNA
<213> Homo sapiens
<400> 108
                                                                       60
agattggatt cgtagattaa acttgagaaa caaaccataa aagtggaagg ccctctttaa
                                                                       62
ca
<210> 109
<211> 86
<212> DNA
<213> Homo sapiens
<400> 109
gagatggggg tctcgctgtg ttgcccaggc tggtcttgga ctcaagcaat ctgcctgtct
                                                                       60
                                                                       86
cagcctacca aaatgctgga ttatag
<210> 110
<211> 116
<212> DNA
 <213> Homo sapiens
 <400> 110
                                                                       60
 gctaaagcag tcctcctgag tagttaggac tacagacata cacgtgccac cgcgcccagc
                                                                      116
 teegtgttet etttgtttee etgeeteetg etetteeact tatetttgea tggeag
 <210> 111
 <211> 45
 <212> DNA
 <213> Homo sapiens
                                                                        45
 ggaaagacga tggagggcag cgaagagagt caccgaccat gacgg
 <210> 112
 <211> 5138
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> misc feature
 <222> 31..33
 <223> ATG
 <221> misc_feature
```

```
<222> 262..264
<223> TAG
<221> polyA_signal
<222> 5111..5116
<223> AATAAA
<400> 112
ctgctgtccc tggtgctcca cacgtactcc atg cgc tac ctg ctc ccc agc gtc
                                                                     54
                                Met Arg Tyr Leu Leu Pro Ser Val
                                1
gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg
                                                                    102
Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
                                           20
                       15
egg etg etc tee gee tte etg eec ege tte tae caa geg etg gae
                                                                    150
Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
                                       35
                    30
gac egg etg tae tge gte tae eag age atg gtg etc tte tte tte gag
                                                                    198
Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Glu
                                   50
                45
aat tac acc ggg gtc cag ttg act gga ttg ttg ctg aca tct tgg cca
                                                                    246
Asn Tyr Thr Gly Val Gln Leu Thr Gly Leu Leu Leu Thr Ser Trp Pro
                               65
            60
                                                                    294
tca ggc aga atg cgc tag gacatgtgcg ctacgtgctg aaagaagggt
Ser Gly Arg Met Arg '
        75
taaaatggct gccattgtat gggtgttact ttgctcagca tggaggaatc tatgtaaagc
                                                                    354
gcagtgccaa atttaacgag aaagagatgc gaaacaagtt gcagagctac gtggacgcag
                                                                    414
gaactccaat gtatcttgtg atttttccag aaggtacaag gtataatcca gagcaaacaa
                                                                    474
aagteettte agetagteag geatttgetg eecaaegtgg eettgeagta ttaaaacatg
                                                                    534
                                                                    594
tgctaacacc acgaataaag gcaactcacg ttgcttttga ttgcatgaag aattatttag
atgcaattta tgatgttacg gtggtttatg aagggaaaga cgatggaggg cagcgaagag
                                                                    654
agtcaccgac catgacggaa tttctctgca aagaatgtcc aaaaattcat attcacattg
                                                                    714
atcgtatcga caaaaaagat gtcccagaag aacaagaaca tatgagaaga tggctgcatg
                                                                    774
aacgtttcga aatcaaagat aagatgctta tagaatttta tgagtcacca gatccagaaa
                                                                    834
gaagaaaaag atttcctggg aaaagtgtta attccaaatt aagtatcaag aagactttac
                                                                    894
catcaatgtt gatcttaagt ggtttgactg caggcatgct tatgaccgat gctggaagga
                                                                    954
                                                                    1014
agctgtatgt gaacacctgg atatatggaa ccctacttgg ctgcctgtgg gttactatta
aagcatagac aagtagctgt ctccagacag tgggatgtgc tacattgtct atttttggcg
                                                                   1074
gctgcacatg acatcaaatt gtttcctgaa tttattaagg agtgtaaata aagccttgtt
                                                                    1134
gattgaagat tggataatag aatttgtgac gaaagctgat atgcaatggt cttgggcaaa
                                                                   1194
catacctggt tgtacaactt tagcatcggg gctgctggaa gggtaaaagc taaatggagt
ttctcctgct ctgtccattt cctatgaact aatgacaact tgagaaggct gggaggattg
                                                                    1314
 tgtattttgc aagtcagatg gctgcatttt tgagcattaa tttgcagcgt atttcacttt
                                                                    1374
 ttctgttatt ttcaatttat tacaacttga cagctccaag ctcttattac taaagtattt
                                                                    1434
 agtatettge agetagttaa tattteatet tttgettatt tetacaagte agtgaaataa
                                                                    1494
 attgtattta ggaagtgtca ggatgttcaa aggaaagggt aaaaagtgtt catggggaaa
                                                                    1554
 aagctctgtt tagcacatga ttttattgta ttgcgttatt agctgatttt actcatttta
                                                                    1614
 tatttgcaaa ataaatttct aatatttatt gaaattgctt aatttgcaca ccctgtacac
                                                                    1674
 acagaaaatg gtataaaata tgagaacgaa gtttaaaatt gtgactctga ttcattatag
                                                                    1734
 cagaacttta aatttcccag ctttttgaag atttaagcta cgctattagt acttcccttt
                                                                    1794
 1854
 tcaaaagagt cggtgtgaac cttggttgga ccccaagttc acaagatttt taaggtgatg
                                                                    1914
 agagectgea gacattetge ctagatttae tagegtgtge ettttgeetg ettetetttg
                                                                    1974
 atttcacaga atattcattc agaagtcgcg tttctgtagt gtggtggatt cccactgggc
                                                                    2034
 tetggteett ceettggate eegteagtgg tgetgeteag eggettgeae gtagaettge
                                                                    2094
 taggaagaaa tgcagagcca gcctgtgctg cccactttca gagttgaact ctttaagccc
                                                                    2154
 ttgtgagtgg gcttcaccag ctactgcaga ggcattttgc atttgtctgt gtcaagaagt
                                                                    2214
 tcaccttctc aagccagtga aatacagact taattcgtca tgactgaacg aatttgttta
                                                                    2274
 tttcccatta ggtttagtgg agctacacat taatatgtat cgccttagag caagagctgt
                                                                    2334
 gttccaggaa ccagatcacg atttttagcc atggaacaat atatcccatg ggagaagacc
                                                                    2394
 tttcagtgtg aactgttcta tttttgtgtt ataatttaaa cttcgatttc ctcatagtcc
                                                                    2454
 tttaagttga catttctgct tactgctact ggatttttgc tgcagaaata tatcagtggc
                                                                    2514
```

2574

```
ccacattaaa cataccagtt ggatcatgat aagcaaaatg aaagaaataa tgattaaggg
aaaattaagt gactgtgtta cactgcttct cccatgccag agaataaact ctttcaagca
                                                                    2634
tcatctttga agagtcgtgt ggtgtgaatt ggtttgtgta cattagaatg tatgcacaca
                                                                    2694
tccatggaca ctcaggatat agttggccta ataatcgggg catgggtaaa acttatgaaa
                                                                    2754
atttcctcat gctgaattgt aattttctct tacctgtaaa gtaaaattta gatcaattcc
                                                                    2814
atgtctttgt taagtacagg gatttaatat attttgaata taatgggtat gttctaaatt
                                                                    2874
tgaactttga gaggcaatac tyttggaatt atgtggattc taactcattt taacaaggta
                                                                    2934
geotgacetg cataagatea setgaatgtt aggttteata gaactataet aatettetea
                                                                    2994
caaaaggtct ataaaataca gtcgttgaaa aaaattttgt atcaaaatgt ttggaaaatt
                                                                    3054
agaagettet cettaacetg tattgatact gaettgaatt attttetaaa attaagagee
                                                                    3114
gratacetae etgraagtet tricacatat cattraaact trigtitgta trattactga
                                                                    3174
tttacagett agttattaat ttttctttat aagaatgeeg tegatgegea tgettttatg
                                                                    3234
tttttcagaa aagggtgtgt ttggatgaaa gtaaaaaaaa aaataaaatc tttcactgtc
tctaatggct gtgctgttta acattttttg accctaaaat tcaccaacag tctcccagta
                                                                    3354
cataaaatag gottaatgac tggccctgca ttcttcacaa tatttttccc taagctttga
                                                                    3414
gcaaagtttt aaaaaaatac actaaaataa tcaaaactgt taagcagtat attagtttgg
                                                                    3474
ttatataaat tcatctgcaa tttataagat gcatggccga tgttaatttg cttggcaatt
                                                                    3534
ctgtaatcat taagtgatct cagtgaaaca tgtcaaatgc cttaaattaa ctaagttggt
                                                                    3594
gaataaaagt googatotgg otaactotta caccatacat actgatagtt tttcatatgt
                                                                    3654
ttcatttcca tgtgattttt aaaatttaga gtggcaacaa ttttgcttaa tatgggttac
                                                                     3714
ataagettta tttttcctt tgttcataat tatattcttt gaataggtct gtgtcaatca
                                                                     3774
agtgatctaa ctagactgat catagataga aggaaataag gccaagttca agaccagcct
                                                                    3834
gggcaacata tcgagaacct gtctacaaaa aaattaaaaa aaattagcca ggcatggtgg
                                                                    3894
cgtacactga gtagtttgtc ccagctactc gggagggtga ggtgggagga tcgcttcagc
                                                                     3954
ccaggaggtt gagattgcag tgagccatgg acataccact gcactacage ctaggtaaca
gcacgagacc ccaactetta gaaaatgaaa aggaaatata gaaatataaa atttgettat
                                                                     4074
tatagacaca cagtaactcc cagatatgta ccacaaaaaa tgtgaaaaga gagagaaatg
                                                                     4134
tctaccaaag cagtattttg tgtgtataat tgcaagcgca tagtaaaata attttaacct
                                                                     4194
taatttgttt tragtagtgt tragattgaa gattgagtga aatattttct tggcagatat
                                                                     4254
tecgtatetg gtggaaaget acaatgeaat gtegttgtag ttttgcatgg ettgetttat
                                                                     4314
aaacaagatt ttttctccct ccttttgggc cagttttcat tacgagtaac tcacactttt
                                                                     4374
 tgattaaaga acttgaaatt acgttatcac ttagtataat tgacattata tagagactat
                                                                     4434
gtaacatgca atcattagaa tcaaaattag tactttggtc aaaatattta caacattcac
                                                                     4494
atacttgtca aatattcatg taattaactg aatttaaaac cttcaactat tatgaagtgc
                                                                     4554
 tegtetgtae aategetaat ttacteagtt tagagtaget acaactette gatactatea
                                                                     4614
 tcaatatttg acatcttttc caatttgtgt atgaaaagta aatctattcc tgtagcaact
                                                                     4674
 ggggagtcat atatgaggtc aaagacatat accttgttat tataatatgt atactataat
                                                                     4734
 aatagctggt tatcctgagc aggggaaaag gttattttta ggaaaaccac ttcaaataga
                                                                     4794
 aagctgaagt acttctaata tactgaggga agtataatat gtggaacaaa ctctcaacaa
                                                                     4854
 aatgtttatt gatgttgatg aaacagatca gtttttccat ccggattatt attggttcat
                                                                     4914
 gattttatat gtgaatatgt aagatatgtt ctgcaatttt ataaatgttc atgtcttttt
                                                                     4974
 ttaaaaaagg tgctattgaa attctgtgtc tccagcaggc aagaatactt gactaactct
                                                                     5034
 ttttgtctct ttatggtatt ttcagaataa agtctgactt gtgtttttga gattattggt
                                                                     5094
                                                                     5138
 gcctcattaa ttcagcaata aaggaaaata tgcatctcaa aaat
 <210> 113
 <211> 5224
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <222> 31..33
 <223> ATG
 <221> misc_feature
 <222> 262..264
 <223> TAG
 <221> polyA_signal
  <222> 5197..5202
  <223> AATAAA
  <400> 113
  ctgctgtccc tggtgctcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc
                                                                        54
```

Met Arg Tyr Leu Leu Pro Ser Val

								1				5				
			~~~	200	~~~	ccc	200		ata	tta	acc		aaa	atc	taa	102
grg	CEC	ctg	Glv	whr	Δla	Pro	Thr	Tvr	Val	Leu	Ala	Tro	Glv	Val	Trp	
vaı	10	ьец	GIY	1111	ліа	15	1111	-1-	,	200	20		1			
caa	cta	ctc	trc	acc	ttc	ctg	ccc	acc	cac	ttc	tac	caa	gcg	ctg	gac	150
Ara	T.eu	Len	Ser	Ala	Phe	Leu	Pro	Ala	Arq	Phe	Tyr	Gln	Ala	Leu	Asp	
25	LCu	200			30				•	35	-				40	
gac	caa	cta	tac	tac	qtc	tac	cag	agc	atg	gtg	ctc	ttc	ttc	ttc	gag	198
Asp	Ara	Leu	Tvr	Cys	Val	Tyr	Gln	Ser	Met	Val	Leu	Phe	Phe	Phe	Glu	
			ļ	45					50					55		
aat	tac	acc	ggg	gtc	cag	ttg	act	gga	ttg	ttg	ctg	aca	tct	tgg	cca	246
Asn	Tyr	Thr	Gly	Val	Gln	Leu	Thr	Gly	Leu	Leu	Leu	Thr	Ser	Trp	Pro	
			60					65					70			
tca	ggc	aga	atg	cgc	tag	gaca	atgto	gcg	ctac	gtgct	g a	aaga	aggg	t		294
Ser	Gly	Arg	Met	Arg	*											
		75														254
taaa	aatg	gct	gcca	ttgt	at g	ggtg	ttact	t tt	gctc	agga	gat	gggg	gtc	tcgc	tgtgtt	354
gcc	cagg	ctg	gtct	tgga	ct c	aagc	aatct	gc.	ctgt	ctca	gcc	tacc	aaa	atgc	tggatt	414
ata	gcat	gga	ggaa	tcta	tg t	aaag	cgca	g tg	ccaa	attt	aac	gaga	aag	agat	gcgaaa	474
caa	gttg	cag	agct	acgt	gg a	cgca	ggaa	c tc	caat	gtat	ctt	gtga	ttt	LLCC	agaagg	534
tac	aagg	tat	aatc	caga	gc a	aaca	aaagi	t cc	tttc	agct	agt	cagg	cat	ttgc	tgccca	594
acg	tggc	ctt	gcag	tatt	aa a	acat	gtgc	t aa	cacc	acga	ata	aagg	caa	ctca	cgttgc	654
ttt	tgat	tgc	atga	agaa	tt a	ittta	gatg	c aa	ttta	tgat	gtt	acgg	tgg	ttta	tgaagg	714
gaa	agac	gat	ggag	ggca	gc g	jaaga	gagt	cac	cgac	catg	acg	gaat	ttc	tctg	caaaga	774
atg	tcca	aaa	attc	atat	tc a	catt	gatc	g ta	tcga	caaa	aaa	gatg	TCC	caga	agaaca	834
aga	acat	atg	agaa	gatg	gc t	gcat	gaac	g tt	tcga	aatc	aaa	gata	aga	tgct	tataga	894
att	ttat	gag	tcac	caga	tc c	agaa	agaa	g aa	aaag	attt	cct	ggga	aaa	gtgt	taattc	954
caa	atta	agt	atca	agaa	ga c	ttta	ccat	c aa	tgtt	gatc	tta	agtg	gtt	tgac	tgcagg	1014 1074
cat	gctt	atg	accg	atgc	tg g	jaagg	aagc	t gt	atgt	gaac	acc	tgga	tat	acgg	aaccct	1134
act	tggc	tgc	ctgt	gggt	ta c	ctatt	aaag	c at	agac	aagt	agc	tgtc	FCC	agac	agtggg	1194
atg	tgct	aca	ttgt	ctat	tt t	tggc	ggct	g ca	catg	acat	caa	attg	200	tata	aattta	1254
tta	agga	gtg	taaa	taaa	gc c	cttgt	tgat	t ga	agat	tgga	taa	caya	acc	cycy	acgaaa	1314
gct	gata	tgc	aarg	gtct	rg g	ggcaa	acat	a cc	tage	cgca	caa	+++	ayu	tras	gggctg	1374
ctg	gaag	ggt	aaaa	gcta	aa t	ggag	TEEC	T CC	tget	.c.g.	CCa	2+44	cta	catt	ctaatg	1434
aca	actt	gag	aagg	ctgg	ga g	ggatt	gtgt	a	.ccgc	tta	cay	tatt	202	actt	tttgag gacagc	1494
cat	taat	ttg	cago	gtat	יבר כ	eactt	tagt	2 to	ttacc	eact	act	taat	att	tcat	cttttg	1554
tcc	aagc	tct	cacc	.acta	iaa y	guauu saata	aatt	a te	ttta	uddaa	ata	tcac	gat	atto	aaagga	1614
CLL	accc	Cta	caag	ttes	ta a	adata agas	aaacc	e te	tatt	tage	aca	tgat	ttt	atto	tattgc	1674
aag	ggta	aaa	ayty	++=	te s	39944 attt	atat	t to	rcaaa	ataa	att	tcta	ata	ttta	ttgaaa	1734
944	atta atta	att	tace	caco	oct o	rtaca	caca	o aa	aato	rotat	aaa	atat	gag	aaco	aagttt	1794
222	2++0	rtra	ctct	ratt	ca	rtata	acaa	a ac	ttta	aatt	tcc	cago	ttt	ttqa	agattt	1854
220	rctac	ract	atta	artac	·++	coctt	tate	t at	acca	taaq	tac	ttga	aaa	cgtt	aaggtt	1914
ttc	tatt	tta	tttt	att	itt i	ttaat	atca	a aa	agagt	caat	gto	aaco	ttg	gttg	gacccc	1974
aac	ittca	caa	gatt	ttta	aag (	gtgat	gaga	g c	ctgca	agaca	tto	tgcc	tag	attt	actage	2034
ato	rtacc	ttt	tace	tact	tc	tcttt	gatt	t ca	acaga	aatat	tca	ttca	agaa	gtcg	gcgtttc	2094
tat	agto	rt.aa	taga	attc	cca	ctaac	ctct	g gt	cctt	ccct	tgg	gatco	cgt	cagt	ggtgct	2154
act	cago	aac	tta	cacq	tag	actto	rctag	g a	agaaa	atgca	gag	ccag	gcct	gtg	ctgccca	2214
ctt	tcac	ragt.	t.gaa	actc	ttt :	aagco	ctto	it ga	agtgg	ggctt	cad	cago	ctac	tgca	agaggca	2274
ttt	taca	attt	atc	tatai	tca	agaag	gttca	ic ci	ttcto	caago	cag	gtgaa	aata	caga	acttaat	2334
t.co	rtcat	cac	tgaa	acqaa	att	tattt	cattt	c c	catta	aggtt	: tag	gtgga	agct	acad	cattaat	2394
ato	rtato	cacc	ttad	rage	aaq	ageto	gtgtt	c c	aggaa	accag	, ato	cacga	attt	ttag	gccatgg	2454
aad	caata	atat	ccca	ataa	gag	aagad	ccttt	c a	gtgt	gaact	gt!	tctai	tttt	tgt	gttataa	2514
tti	taaad	ette	gati	ttcc	tca	tagto	ccttt	aa	gttga	acatt	: tc	tgcti	tact	gcta	actggat	2574
titi	ttaci	taca	gaa	atat	atc	agtg	gccca	ac a	ttaa	acata	CC	agtt	ggat	cate	gataagc	2634
aa	aato	aaaq	aaa	taat	gat	taagg	ggaaa	aa t	taag	tgact	gt:	gtta	cact	gct.	tctccca	2694
ta	ccaa	agaa	taa	actc	ttt	caage	catca	at c	tttg	aagag	y to	gtgt	ggtg	tga	attggtt	2754
t.a	tata	catt	aga	atat	atq	caca	catco	ca t	ggac	actca	a gg	atat	agtt	ggc	ctaataa	2814
tc	aaaa	catq	aat	aaaa	ctt	atga	aaati	tt c	ctca	tgct	g aa	ttgt	aatt	ttc	tcttacc	2874
tg	taaa	gtaa	aat	ttag	atc	aatt	ccat	gt c	tttg	ttaa	g ta	cagg	gatt	taa	tatattt	2934
tg	aata	taat	ggg	tatg	ttc	taaa	tttg	aa c	tttg	agag	g ca	atac	tgtt	gga	attatgt	2994

```
ggattctaac tcattttaac aaggtagcct gacctgcata agatcacttg aatgttaggt
                                                                    3054
ttcatagaac tatactaatc ttctcacaaa aggtctataa aatacagtcg ttgaaaaaaa
                                                                    3114
ttttgtatca aaatgtttgg aaaattagaa gcttctcctt aacctgtatt gatactgact
                                                                    3174
tgaattattt totaaaatta agagoogtat acctacotgt aagtotttto acatatoatt
                                                                    3234
taaacttttg tttgtattat tactgattta cagcttagtt attaattttt ctttataaga
atgccgtcga tgtgcatgct tttatgtttt tcagaaaagg gtgtgtttgg atgaaagtaa
                                                                    3354
aaaaaaaaat aaaatctttc actgtctcta atggctgtgc tgtttaacat tttttgaccc
                                                                    3414
taaaattcac caacagtctc ccagtacata aaataggctt aatgactggc cctgcattct
                                                                    3474
tcacaatatt tttccctaag ctttgagcaa agttttaaaa aaatacacta aaataatcaa
                                                                    3534
aactgttaag cagtatatta gtttggttat ataaattcat ctgcaattta taagatgcat
                                                                    3594
                                                                     3654
ggccgatgtt aatttgcttg gcaattctgt aatcattaag tgatctcagt gaaacatgtc
aaatgcctta aattaactaa gttggtgaat aaaagtgccg atctggctaa ctcttacacc
                                                                     3714
atacatactg atagtttttc atatgtttca tttccatgtg atttttaaaa tttagagtgg
                                                                     3774
caacaatttt gcttaatatg ggttacataa gctttatttt ttcctttgtt cataattata
                                                                     3834
                                                                     3894
ttctttgaat aggtctgtgt caatcaagtg atctaactag actgatcata gatagaagga
                                                                     3954
aataaggcca agttcaagac cagcctgggc aacatatcga gaacctgtct acaaaaaaat
taaaaaaat tagccaggca tggtggcgta cactgagtag tttgtcccag ctactcggga
                                                                     4014
gggtgaggtg ggaggatcgc ttcagcccag gaggttgaga ttgcagtgag ccatggacat
                                                                     4074
                                                                     4134
accactgcac tacagcctag gtaacagcac gagaccccaa ctcttagaaa atgaaaagga
aatatagaaa tataaaattt gottattata gacacacagt aactoocaga tatgtaccac
                                                                     4194
aaaaaatgtg aaaagagaga gaaatgtcta ccaaagcagt attttgtgtg tataattgca
                                                                     4254
agggcatagt aaaataattt taaccttaat ttgtttttag tagtgtttag attgaagatt
                                                                     4314
gagtgaaata ttttcttggc agatattccg tatctggtgg aaagctacaa tgcaatgtcg
                                                                     4374
ttgtagtttt gcatggcttg ctttataaac aagatttttt ctccctcctt ttgggccagt
                                                                     4434
tttcattacg agtaactcac actttttgat taaagaactt gaaattacgt tatcacttag
                                                                     4494
tataattgac attatataga gactatgtaa catgcaatca ttagaatcaa aattagtact
                                                                     4554
ttggtcaaaa tatttacaac attcacatac ttgtcaaata ttcatgtaat taactgaatt
                                                                     4614
                                                                     4674
taaaaccttc aactattatg aagtgctcgt ctgtacaatc gctaatttac tcagtttaga
gtagctacaa ctcttcgata ctatcatcaa tatttgacat cttttccaat ttgtgtatga
                                                                     4734
                                                                     4794
aaagtaaatc tattcctgta gcaactgggg agtcatatat gaggtcaaag acatatacct
tgttattata atatgtatac tataataata gctggttatc ctgagcaggg gaaaaggtta
                                                                     4854
tttttaggaa aaccacttca aatagaaagc tgaagtactt ctaatatact gagggaagta
                                                                     4914
taatatgtgg aacaaactct caacaaaatg tttattgatg ttgatgaaac agatcagttt
                                                                     4974
ttccatccgg attattattg gttcatgatt ttatatgtga atatgtaaga tatgttctgc
                                                                     5034
                                                                     5094
aattttataa atgttcatgt cttttttaa aaaaggtgct attgaaattc tgtgtctcca
gcaggcaaga atacttgact aactcttttt gtctctttat ggtattttca gaataaagtc
                                                                     5154
tgacttgtgt ttttgagatt attggtgcct cattaattca gcaataaagg aaaatatgca
                                                                     5214
                                                                     5224
tctcaaaaat
<210> 114
<211> 4863
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
 <222> 31..33
 <223> ATG
 <221> misc_feature
 <222> 745..747
 <223> TAG
 <221> polyA_signal
 <222> 4836..4841
 <223> AATAAA
 <400> 114
 etgetgtece tggtgeteca caegtactee atg ege tae etg etg eee age gte
                                                                       54
                                  Met Arg Tyr Leu Leu Pro Ser Val
                                  1
 gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg
                                                                       102
 Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
                                             20
                         15
                                                                       150
 cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac
 Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
```

25					30					35					40	
gac gac	caa	cta	tac	tac	atc	tac	caq	agc	atg		ctc	ttc	ttc	ttc	gag	198
Asn	Ara	Leu	Tyr	Cvs	Val	Tvr	Gln	Ser	Met	Val	Leu	Phe	Phe	Phe	Glu	
_				45					50					55		
aat	tac	acc	ggg	gtc	cag	cat	gga	gga	atc	tat	gta	aag	cgc	agt	gcc	246
Asn	Tyr	Thr	Gly	Val	Gln	His	Gly	Gly	Ile	Tyr	Val	Lys	Arg	Ser	Ala	
			60					65					70			004
aaa	ttt	aac	gag	aaa	gag	atg	cga	aac	aag	ttg	cag	agc	tac	gtg	gac	294
Lys	Phe	Asn	Glu	Lys	Glu	Met	Arg	Asn	Lys	Leu	Gln	Ser	Tyr	Val	Asp	
		75					80					85				2.40
gca	gga	act	cca	atg	tat	ctt	gtg	att	ttt	cca	gaa	ggt	aca	agg	tat	342
Ala	Gly	Thr	Pro	Met	Tyr		Val	Ile	Phe	Pro	Glu	GТĀ	Thr	Arg	Tyr	
	90					95					100					390
aat	cca	gag	caa	aca	aaa	gtc	ctt	tca	gct	agt	cag	gca	777	gct	310	390
Asn	Pro	Glu	Gln	Thr		Val	Leu	Ser	Ala		Gin	Aia	Fne	Ala	120	
105					110					115					120	420
caa	cgt	gaa	t,tt	ctc	tgc	aaa	gaa	tgt	cca	aaa	att	cat	att	cac	att	438
Gln	Arg	Glu	Phe	Leu	Cys	Lys	Glu	Cys	Pro	ьуs	TTE	HIS	TIE	125	iie	
			:	125					130					135		486
gat	cgt	atc	gac	aaa	aaa	gat	gtc	cca	gaa	gaa	caa	gaa	cat	atg	aga	400
Asp	Arg	Ile	Asp	Ьуs	Lys	Asp	Val	Pro	GIu	GIu	GIN	GIU	nis	Met	AIG	
			140					145					150			534
aga	tgg	ctg	cat	gaa	cgt	ttc	gaa	atc	aaa	gat	aag	atg	CLL	ata	gaa	534
Arg	Trp	Leu	His	Glu	Arg	Phe		Ile	Lys	Asp	Lys	Met	Leu	TTE	GIU	
		155					160					165				E00
ttt	tat	gag	tca	cca	gat	cca	gaa	aga	aga	aaa	aga	ttt	CCT	ggg	aaa	582
Phe	Tyr	Glu	Ser	Pro	Asp			Arg	Arg	Lys	Arg	Pne	Pro	GIŞ	ьуs	
	170					175					180					630
agt	gtt	aat	tcc	aaa	tta	agt	atc	aag	aag	act	tta	cca	tca	atg	ttg	630
Ser	Val	Asn	Ser	Lys			Ile	Lys	Lys	Thr	Leu	Pro	Ser	met	ьеи	
185					190					195				~~~	200	678
atc	tta	agt	ggt	ttg	act	gca	ggc	atg	ctt	arg	acc	gat	gci	gya	agg	070
Ile	Leu	Ser	Gly			Ala	GIY	Met			rnr	Asp	Ald	215	Arg	
				205	•				210			_++	~~~			726
aag	ctg	tat	gtg	aac	acc	tgg	ata	tat	. gga	acc	Cta	. CLL	Clu	Cve	ctg	720
Lys	Leu	Туг			Thr	rrp	) TTE			THE	гeп	пец	230	Cys	Leu	
			220					225		~+ ~+					+	777
			att				aca	lagica	iget	gici	.ccay	ac c	gcgg	gucg		
Trp	Val		: Ile	Lys	: Ala	. *										
		235				<b></b>		+.	*2 <i>4</i> 25		. ++0	.+++	cta	aatt	tattaa	837
gct	acat	tgt	ctat	יבבבו	gg c	ggct	gcac	ia ug	taca	.caac	. ans	)	ata.	acga	tattaa	897
gga	gtgt	aaa	taaa	igcet	tg t	tgat	.cgac	19 al	-tggc	caac	. aya	2000	etca	aaaa	aagctg	957
ata	tgca	atg	gtct	cggg	gca e	lacal	.acct	-9 91	ctat	ccat		ctat	raa.	ctaa	tgctgg	1017
aag	ggta	laaa	gcita	aat	gga s	-~+~+		.y c	raant	cana	tac	rcta	att	tttt	gagcatt	1077
CCC	gaga	agg	ctgg	ggagg	gat i	.g.cg.	-acci	y	tttc	ettt	a cgs	acas	actt	gaca	agctcca	1137
aat	ttg	cagc	gtai	CLEC	act t			- t ~	react	tacti	- aat	atti	cat	cttt	agctcca stgctta	1197
ago	CCCI	catt	acte	aaay	lat i	- cay	rate	+:	anna	aatai	t cac	raato	rttc	aaaa	gaaagg	1257
בבנ	cta	caag	CCAS	gryad	aat o	2222	retei	ta t	ttam	racai	t dat	tttt	atta	tatt	gcgtta	1317
gta	laaa	agug		acgg	gya d		+++~	ra a	aata:	aatti	t cta	aatai	ttta	tta	aaattgc	1377
ככנ	igct	yact Lace	LCA	cuca	tac '	cata	araa	aa t	aata	taaa	a tai	tgag	aaco	aagi	ttaaaa	1437
C C 2	aact'	ugca atat	cac	test	tat (	acac	ayaa.		aaati	ttcc	c ag	cttt	ttaa	aga	ttaagc	1497
E C (	ucga.	2++2	yat	cttc	cat (	ttat	ctat	מכ כ	ataa	atac.	t ta	aaaa	catt	aaq	gttttct	1557
(a)	-yct	a ∟ i d + + + +	ged	++++	taa .	tato	2232	മരസ	teaa	tata	a ac	ctta	atta	gac	cccaagt	1617
gen	2022	~	4++	2244	tra	tasa	agee	=- 4	agac	atto	t ac	ctaq	attt	act	agcgtgt	1677
~~	atad	yaıl tass	+44	4499 ++4+	ctt	tast	ttca	ca d	aata	ttca	t to	agaa	gtca	cat	ttctgta	1737
90	ataa	tare	tta	CCSC	taa	acta	taat	cc t	tece	ttaa	a tc	ccat	cagt	ggt	gctgctc	1797
90	guga	ttaa	200	tage	-22 -23	acta	ggaa	ga a	atoc	agag	c ca	gcct	gtgc	tgc	ccacttt	1857
ag	aaa+	tass	ctc	ttta	age	cctt	ataa	at a	gact	tcac	c aq	ctac	tgca	gag	gcatttt	1917
αc	attt	atct	: ata	tcaa	gaa	atto	acct	tc t	caag	ccag	t ga	aata	caga	ctt	aattcgt	1977
CA	tgac	toaa	CGA	attt	att	tatt	tccc	at t	aggt	ttag	t gg	agct	acac	att	aatatgt	2037
		- 5 0			_					_						

```
ategeettag ageaagaget gtgtteeagg aaceagatea egatttttag eeatggaaca
                                                                    2097
atatatccca tgggagaaga cctttcagtg tgaactgttc tatttttgtg ttataattta
                                                                    2157
aacttcgatt tcctcatagt cctttaagtt gacatttctg cttactgcta ctggattttt
                                                                    2217
gctgcagaaa tatatcagtg gcccacatta aacataccag ttggatcatg ataagcaaaa
                                                                    2277
tgaaagaaat aatgattaag ggaaaattaa gtgactgtgt tacactgctt ctcccatgcc
                                                                    2337
agagaataaa ctctttcaag catcatcttt gaagagtcgt gtggtgtgaa ttggtttgtg
                                                                    2397
tacattagaa tgtatgcaca catccatgga cactcaggat atagttggcc taataatcgg
                                                                    2457
ggcatgggta aaacttatga aaatttcctc atgctgaatt gtaattttct cttacctgta
                                                                    2517
aagtaaaatt tagatcaatt ccatgtcttt gttaagtaca gggatttaat atattttgaa
                                                                   2577
tataatgggt atgttctaaa tttgaacttt gagaggcaat actgttggaa ttatgtggat
tctaactcat tttaacaagg tagcctgacc tgcataagat cacttgaatg ttaggtttca
                                                                    2697
tagaactata ctaatcttct cacaaaaggt ctataaaata cagtcgttga aaaaaatttt
                                                                    2757
gtatcaaaat gtttggaaaa ttagaagctt ctccttaacc tgtattgata ctgacttgaa
                                                                    2817
                                                                    2877
ttattttcta aaattaagag ccgtatacct acctgtaagt cttttcacat atcatttaaa
cttttgtttg tattattact gatttacagc ttagttatta atttttcttt ataagaatgc
                                                                    2937
cgtcgatgtg catgctttta tgtttttcag aaaagggtgt gtttggatga aagtaaaaa
                                                                    2997
aaaaataaaa totttoactg tototaatgg otgtgotgtt taacattttt tgaccotaaa
attcaccaac agtctcccag tacataaaat aggcttaatg actggccctg cattcttcac
                                                                   3117
aatatttttc cctaagcttt gagcaaagtt ttaaaaaaaat acactaaaat aatcaaaact
                                                                    3177
gttaagcagt atattagttt ggttatataa attcatctgc aatttataag atgcatggcc
                                                                    3237
gatgttaatt tgcttggcaa ttctgtaatc attaagtgat ctcagtgaaa catgtcaaat
                                                                    3297
gccttaaatt aactaagttg gtgaataaaa gtgccgatct ggctaactct tacaccatac
                                                                    3357
atactgatag tttttcatat gtttcatttc catgtgattt ttaaaattta gagtggcaac
                                                                    3417
aattttgctt aatatgggtt acataagctt tattttttcc tttgttcata attatattct
                                                                    3477
ttgaataggt ctgtgtcaat caagtgatct aactagactg atcatagata gaaggaaata
                                                                    3537
aggccaagtt caagaccagc ctgggcaaca tatcgagaac ctgtctacaa aaaaattaaa
                                                                    3597
aaaaattagc caggcatggt ggcgtacact gagtagtttg tcccagctac tcgggagggt
                                                                    3657
gaggtgggag gatcgcttca gcccaggagg ttgagattgc agtgagccat ggacatacca
                                                                    3717
                                                                    3777
ctgcactaca gcctaggtaa cagcacgaga ccccaactct tagaaaatga aaaggaaata
tagaaatata aaatttgctt attatagaca cacagtaact cccagatatg taccacaaaa
                                                                    3837
aatgtgaaaa gagagagaaa tgtctaccaa agcagtattt tgtgtgtata attgcaagcg
                                                                    3897
catagtaaaa taattttaac cttaatttgt ttttagtagt gtttagattg aagattgagt
                                                                    3957
gaaatatttt cttggcagat attccgtatc tggtggaaag ctacaatgca atgtcgttgt
                                                                    4017
agttttgcat ggcttgcttt ataaacaaga ttttttctcc ctccttttgg gccagttttc
                                                                    4077
attacgagta actcacactt tttgattaaa gaacttgaaa ttacgttatc acttagtata
                                                                     4137
attgacatta tapagagact atgtaacatg caatcattag aatcaaaatt agtactttgg
                                                                     4197
tcaaaatatt tacaacattc acatacttgt caaatattca tgtaattaac tgaatttaaa
                                                                     4257
accttcaact attatgaagt gctcgtctgt acaatcgcta atttactcag titagagtag
                                                                     4317
ctacaactct tcgatactat catcaatatt tgacatcttt tccaatttgt gtatgaaaag
                                                                    4377
taaatctatt cctgtagcaa ctggggagtc atatatgagg tcaaagacat ataccttgtt
attataatat gtatactata ataatagctg gttatcctga gcaggggaaa aggttatttt
                                                                     4497
taggaaaacc acttcaaata gaaagctgaa gtacttctaa tatactgagg gaagtataat
                                                                     4557
atgtggaaca aactctcaac aaaatgttta ttgatgttga tgaaacagat cagtttttcc
                                                                     4617
atccggatta trattggttc atgattttat atgtgaatat gtaagatatg ttctgcaatt
                                                                     4677
 ttataaatgt tcatgtcttt ttttaaaaaa ggtgctattg aaattctgtg tctccagcag
                                                                    4737
 gcaagaatac ttgactaact ctttttgtct ctttatggta ttttcagaat aaagtctgac
                                                                    4797
 ttgtgttttt gagattattg gtgcctcatt aattcagcaa taaaggaaaa tatgcatctc
                                                                     4857
                                                                     4863
 aaaaat
 <210> 115
 <211> 5022
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <222> 31..33
 <223> ATG
 <221> misc_feature
 <222> 904..906
 <223> TAG
 <221> polyA_signal
```

<222> 4995..5000

	> AA															
<400 ctgc	> 11 tgtc	s cc t	ggtg	ctcc	a ca	cgta	ctcc	Met	cgc Arg	tac Tyr	ctg Leu	ctg Leu 5	ccc Pro	agc Ser	gtc Val	54
gtg Val	Leu	ctg Leu	ggc Gly	acg . Thr .	Ala	Pro '	acc Thr	l tac Tyr	gtg Val	Leu	gcc Ala 20	tgg Trp	ggg ggg	gtc Val	tgg Trp	102
Arg	Leu	Leu	tcc Ser	Ala	ttc Phe 30	Leu	Pro	Ala	Arg	ttc Phe 35	tac Tyr	GIN	Ala	ьeu	40	150
gac Asp	Arg	Leu	tac Tyr	tgc Cys 45	gtc Val	Tyr	Gln	Ser	мет 50	vaı	Leu	Pne	FIIE	55	GIU	198
Asn	Tyr	Thr	Gly 60	gtc Val	Gln	His	Gly	65 GIY	IIe	туr	Val	гуѕ	70	Ser	VIG	246
aaa Lys	ttt Phe	aac Asn 75	gag Glu	aaa Lys	gag Glu	atg Met	cga Arg 80	aac Asn	aag Lys	ttg Leu	cag Gln	agc Ser 85	tac Tyr	gtg Val	gac Asp	294
Ala	Gly	act Thr	cca Pro	Met	Tyr	Leu 95	Val	Ile	Phe	Pro	100	GIĀ	THE	AIG	ığı	342
aat Asn 105	CCA	gag Glu	caa Gln	aca Thr	aaa Lys 110	atc	ctt Leu	tca Ser	gct Ala	agt Ser 115	cag Gln	gca Ala	ttt Phe	gct Ala	gcc Ala 120	390
<b>~~</b>	cgt Arg	ggc	ctt Leu	gca Ala 125	gta	tta Leu	aaa Lys	His	gtg Val 130	cta Leu	aca Thr	cca Pro	cga Arg	ata Ile 135	aag Lys	438
Ala	Thr	His	gtt Val 140	gct Ala	Phe	Asp	Cys	Met 145	гЛs	Asn	туг	ьец	150	ATG	116	486
Tyr	Asp	Val	acg Thr	Val	Val	Tyr	Glu 160	GIĀ	гĀЗ	Asp	Asp	165	GIY	GIII	nrg	534
aga Arg	gag Glu 170	tca Ser	ccg Pro	acc Thr	atg Met	acg Thr 175	σaa	ttt	ctc Leu	tgc Cys	aaa Lys 180	gaa Glu	tgt Cys	cca Pro	aaa Lys	582
Ile	cat His	att	cac His	Ile	Asp	cgt Arg	Ile	Asp	Lys	Lys 195	Asp	Val	Pro	GIU	200	630
925	~==	cat His	atg Met	aga Arg 205	aga Arg	taa	ctg Leu	cat His	gaa Glu 210	Arg	ttc Phe	gaa Glu	atc Ile	aaa Lys 215	rap	678
Lys	Met	: Le	ata 1 Ile 220	gaa Glu	ttt Phe	Tyr	Glu	Ser 225	Pro	) Asp	Pro	GIU	230	Arg	гуs	726
Ar	g Phe	Pro	ggg Gly	aaa Lys	Ser	. Val	. Asr 240	n Ser )	. Lys	s rec	ı ser	245	erñs	. шуз	act Thr	774
Le	u Pro	a to o Se:	a ato r Met	: Let	ı Ile	255	ı Sei	c GTZ	, re	ı Tnı	260	) v GT7	/ Met	. nec	atg Met	822
ac Th 26	c ga r As	- ~~	t gga a Gly	a agg	g aag g Lys 27	g cto s Lev	tai	t gtg r Va	g aad l Asi	c acc n Thi 27!	TI	g ata p Ile	a tat e Tyi	gga Gly	a acc 7 Thr 280	870
ct	a ct u Le	u Gl	c tgo y Cy:	s Le	g tg u Tr	g gti p Va	l Th	r Ile	e Ly: 29	a gca s Ala	a tag					916
gt tt	ctcc gttt	agac .cctg	agt aat	aaaa	tat	gcta ggag	catt tgta	gt c aa t	tatt aaag	tttg actt	g cg g tt	gctg gatt	caca gaag	tga att	catcaaa ggataat	976 1036

agaatttgtg acgaaagctg atatgcaatg gtcttgggca aacatacctg gttgtacaac 1096 tttagcatcg gggctgctgg aagggtaaaa gctaaatgga gtttctcctg ctctgtccat 1156 ttcctatgaa ctaatgacaa cttgagaagg ctgggaggat tgtgtatttt gcaagtcaga 1216 tggctgcatt tttgagcatt aatttgcagc gtatttcact ttttctgtta ttttcaattt 1276 attacaactt gacageteca agetettatt actaaagtat ttagtatett geagetagtt 1336 aatatttcat cttttgctta tttctacaag tcagtgaaat aaattgtatt taggaagtgt 1396 caggatgttc aaaggaaagg gtaaaaagtg ttcatgggga aaaagctctg tttagcacat gattttattg tattgcgtta ttagctgatt ttactcattt tatatttgca aaataaattt 1456 1516 ctaatattta ttgaaattgc ttaatttgca caccctgtac acacagaaaa tggtataaaa 1576 1636 tatgagaacg aagtttaaaa ttgtgactct gattcattat agcagaactt taaatttccc agetttttga agatttaage taegetatta gtaetteeet ttgtetgtge cataagtget 1696 tgaaaacgtt aaggttttct gttttgtttt gttttttaa tatcaaaaga gtcggtgtga 1756 accttggttg gaccccaagt tcacaagatt tttaaggtga tgagagcctg cagacattct 1816 gcctagattt actagcgtgt gccttttgcc tgcttctctt tgatttcaca gaatattcat 1876 tcagaagtcg cgtttctgta gtgtggtgga ttcccactgg gctctggtcc ttcccttgga 1936 tecegteagt ggtgetgete ageggettge aegtagaett getaggaaga aatgeagage 1996 2056 cagcetgtgc tgcccacttt cagagttgaa ctctttaagc ccttgtgagt gggcttcacc agctactgca gaggcatttt gcatttgtct gtgtcaagaa gttcaccttc tcaagccagt 2116 gaaatacaga cttaattcgt catgactgaa cgaatttgtt tatttcccat taggtttagt 2176 ggagctacac attaatatgt atcgccttag agcaagagct gtgttccagg aaccagatca 2236 cgatttttag ccatggaaca atatatccca tgggagaaga cctttcagtg tgaactgttc 2296 tatttttgtg ttataattta aacttcgatt tcctcatagt cctttaagtt gacatttctg 2356 cttactgcta ctggattttt gctgcagaaa tatatcagtg gcccacatta aacataccag 2416 ttggatcatg ataagcaaaa tgaaagaaat aatgattaag ggaaaattaa gtgactgtgt 2476 tacactgctt ctbccatgcc agagaataaa ctctttcaag catcatcttt gaagagtcgt 2536 2596 gtggtgtgaa ttggtttgtg tacattagaa tgtatgcaca catccatgga cactcaggat atagttggcc taataatcgg ggcatgggta aaacttatga aaatttcctc atgctgaatt 2656 2716 gtaattttct cttacctgta aagtaaaatt tagatcaatt ccatgtcttt gttaagtaca 2776 gggatttaat atattttgaa tataatgggt atgttctaaa tttgaacttt gagaggcaat actgttggaa ttatgtggat tctaactcat tttaacaagg tagcctgacc tgcataagat 2836 cacttgaatg ttaggtttca tagaactata ctaatcttct cacaaaaggt ctataaaata 2896 cagtcgttga aaaaaatttt gtatcaaaat gtttggaaaa ttagaagctt ctccttaacc 2956 tgtattgata ctgacttgaa ttattttcta aaattaagag ccgtatacct acctgtaagt 3016 3076 cttttcacat atcatttaaa cttttgtttg tattattact gatttacagc ttagttatta atttttcttt ataagaatge egtegatgtg catgetttta tgtttttcag aaaagggtgt 3136 gtttggatga aagtaaaaaa aaaaataaaa tctttcactg tctctaatgg ctgtgctgtt 3196 taacattttt tgaccctaaa attcaccaac agtctcccag tacataaaat aggcttaatg actggccctg cattcttcac aatatttttc cctaagcttt gagcaaagtt ttaaaaaaaat 3316 acactaaaat aatcaaaact gttaagcagt atattagttt ggttatataa attcatctgc 3376 3436 aatttataag atgcatggcc gatgttaatt tgcttggcaa ttctgtaatc attaagtgat ctcagtgaaa catgtcaaat gccttaaatt aactaagttg gtgaataaaa gtgccgatct 3496 ggctaactct tacaccatac atactgatag tttttcatat gtttcatttc catgtgattt 3556 ttaaaattta gagtggcaac aattttgctt aatatgggtt acataagctt tatttttcc 3616 tttgttcata attatattct ttgaataggt ctgtgtcaat caagtgatct aactagactg 3676 atcatagata gaaggaaata aggccaagtt caagaccagc ctgggcaaca tatcgagaac 3736 ctgtctacaa aaaaattaaa aaaaattagc caggcatggt ggcgtacact gagtagtttg 3796 teccagetae tegggagggt gaggtgggag gategettea geccaggagg ttgagattge 3856 3916 agtgagccat ggacatacca ctgcactaca gcctaggtaa cagcacgaga ccccaactct tagaaaatga aaaggaaata tagaaatata aaatttgctt attatagaca cacagtaact 3976 cccagatatg taccacaaaa aatgtgaaaa gagagagaaa tgtctaccaa agcagtattt 4036 tgtgtgtata attgcaagcg catagtaaaa taattttaac cttaatttgt ttttagtagt 4096 gtttagattg aagattgagt gaaatatttt cttggcagat attccgtatc tggtggaaag 4156 ctacaatgca atgtcgttgt agttttgcat ggcttgcttt ataaacaaga ttttttctcc 4216 ctccttttgg gccagttttc attacgagta actcacactt tttgattaaa gaacttgaaa 4276 ttacgttatc acttagtata attgacatta tatagagact atgtaacatg caatcattag 4336 aatcaaaatt agtactttgg tcaaaatatt tacaacattc acatacttgt caaatattca 4396 tgtaattaac tgaatttaaa accttcaact attatgaagt gctcgtctgt acaatcgcta 4456 atttactcag tttagagtag ctacaactct tcgatactat catcaatatt tgacatcttt 4516 tccaatttgt gtatgaaaag taaatctatt cctgtagcaa ctggggagtc atatatgagg 4576 tcaaagacat ataccttgtt attataatat gtatactata ataatagctg gttatcctga 4636 gcaggggaaa aggttatttt taggaaaacc acttcaaata gaaagctgaa gtacttctaa 4696

tatactgagg gaagtataat atgtggaaca aactctcaac aaaatgttta ttgatgttgaacagatcagat	g 4876 a 4936
ctgctgtccc tggtgctcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc Met Arg Tyr Leu Leu Pro Ser Val 1 5	
gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp  10 15 20	102
cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp	150
gac cgg ctg tac tgc gtc tac cag agc atg gtg ctc ttc ttc ttc gag Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Glu	198
aat tac acc ggg gtc cag atg tat ctt gtg att ttt cca gaa ggt aca Asn Tyr Thr Gly Val Gln Met Tyr Leu Val Ile Phe Pro Glu Gly Thr	246
agg tat aat cca gag caa aca aaa gtc ctt tca gct agt cag gca ttt Arg Tyr Asn Pro Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe	294
gct gcc caa cgt ggc ctt gca gta tta aaa cat gtg cta aca cca cga Ala Ala Gln Arg Gly Leu Ala Val Leu Lys His Val Leu Thr Pro Arg	342
ata aag gca act cac gtt gct ttt gat tgc atg aag aat tat tta gat Ile Lys Ala Thr His Val Ala Phe Asp Cys Met Lys Asn Tyr Leu Asp	390
gca att tat gat gtt acg gtg gtt tat gaa ggg aaa gac gat gga ggg Ala Ile Tyr Asp Val Thr Val Val Tyr Glu Gly Lys Asp Asp Gly Gly	438
cag cga aga gag tca ccg acc atg acg gaa ttt ctc tgc aaa gaa tgt Gln Arg Arg Glu Ser Pro Thr Met Thr Glu Phe Leu Cys Lys Glu Cys	486
cca aaa att cat att cac att gat cgt atc gac aaa aaa gat gtc cca Pro Lys Ile His Ile His Ile Asp Arg Ile Asp Lys Lys Asp Val Pro	534
gaa gaa caa gaa cat atg aga aga tgg ctg cat gaa cgt ttc gaa atc Glu Glu Glu His Met Arg Arg Trp Leu His Glu Arg Phe Glu Ile	582
aaa gat aag atg ctt ata gaa ttt tat gag tca cca gat cca gaa aga Lys Asp Lys Met Leu Ile Glu Phe Tyr Glu Ser Pro Asp Pro Glu Arg	J
185 190 195 200 aga aaa aga ttt cct ggg aaa agt gtt aat tcc aaa tta agt atc aag	

Arg Lys Arg Phe Pro Gly	Lys Ser Va	l Asn Ser I 210	ys Leu Ser	Ile Lys 215	
205 aag act tta cca tca atg	tta ata tt		to act oca		726
Lys Thr Leu Pro Ser Met	Tou The Le	a age gge o	eu Thr Ala	Glv Met	
	22	54 561 617 1	230		
220 ctt atg acc gat gct gga	and and ct	a tat ata a		ata tat	774
Leu Met Thr Asp Ala Gly	Arg Tye Le	on Tur Val A	an Thr Tro	Ile Tyr	
	240	d lyr var .	245	-	
235 gga acc cta ctt ggc tgc		t act att a			816
gga acc cta ctt ggc tgc	tou Man Va	ol mbr Tle I	ws Ala *		
Gly Thr Leu Leu Gly Cys	255	11 1111 110 1	260		
250 acaagtaget gtetecagae ag	233 ~=~~~			nggctgcaca	876
acaagtaget gtetecagae ag	gigggaige q	geracatege (	rasacctta	rrgattgaag	936
tgacatcaaa ttgtttcctg aa	attiattaa (	gyaytytaaa (	rtcttgggca /	acatacete	996
attggataat agaatttgtg ac	egaaagetg d	acacycaacy y	gctaaatgga :	attteteeta	1056
gttgtacaac tttagcatcg g	ggetgetgg (	aayyytaaaa y	ctacacaga :	ratatattt	1116
ctctgtccat ttcctatgaa c	taatyacaa (	natttacaac	ctggguggut	tttctgtta	1176
gcaagtcaga tggctgcatt t	ttgagcall (	aditiycayt ;	actaaactar	tragtatett	1236
ttttcaattt attacaactt g	acageteca	ttatacasa	tcactcaaat	aaartgtatt	1296
gcagctagtt aatatttcat c	ttttgetta	etacaaaa etacaaaaa	ttcatgggga	aaaagctctg	1356
taggaagtgt caggatgttc a	aaggaaagg	ttaaaaaguy	ttactcattt	tatatttgca	1416
tttagcacat gattttattg t	attgcgtta	ttagotgatt	coacctatec	acacacaaaa	1476
aaataaattt ctaatattta t	tgaaattgc	ttaattigca	cattentat	acacagaaatt	1536
tggtataaaa tatgagaacg a	agtttaaaa	ttgtgactet	gallialiat etasttesst	-tatatata	1596
taaatttccc agcttttga a	gatttaagc	tacgctatta	gtacttccct	tatossasos	1656
cataagtgct tgaaaacgtt a	aggttttct	gttttgttt	gtttttttaa	taccaaaaga	1716
gtcggtgtga accttggttg g	accccaagt	tcacaagatt	tttaaggtga	tastttasas	1776
cagacattct gcctagattt a	ctagcgtgt	gccttttgcc	tgettetett	retetere	1836
gaatattcat tcagaagtcg C	gtttctgta	gtgtggtgga	ttcccactgg	getetggtee	1896
tracettaga tecegteagt g	ratactactc	agcggcttgc	acgtagactt	gctaggaaga	1956
aatgragage cageetgtge t	gcccacttt	cagagttgaa	ctctttaagc	Colligigage	2016
gggcttcacc agctactgca g	raggeatttt	gcatttgtct	gtgtcaagaa	gillaccit	2016
traacreant qaaatacaga C	ttaattcqt	catgactgaa	cgaatttgtt	tatttttttat	2136
taggittagt ggaggtagag a	ttaatatqt	atcgccttag	agcaagagct	gtgttccayy	
aaccadatca cdatttttad C	catogaaca	atatatccca	tgggagaaya	Cottecageg	2196
treactotto tatttttqtq t	tataattta	aacttcgatt	tcctcatagt	CCLLLaagu	2256
gacatttctg cttactgcta c	tagattttt	gctgcagaaa	tatatcagtg	geceacatta	2316
aacataccad tingatcato a	ataagcaaaa	tgaaagaaat	aatgattaag	ggaaaattaa	2376
ataactatat tabactactt 0	steccatgee	agagaataaa	ctctttcaag	Calcalcit	2436
gaagagtogt gtggtgtgaa t	taatttata	tacattagaa	tgtatgcaca	catccatgga	2496
cactcaccat atacttcccc t	taataatcqq	ggcatgggta	aaacttatga	dadittictic	2556
atactasatt ataattttct (	cttacctgta	aagtaaaatt	tagatcaatt	Coalgittit	2616
attaadtaca domatttaat a	atattttgaa	tataatgggt	atgttctaaa	tttgaacttt	2676
gagagggaat actgttggaa 1	ttatqtqqat	tctaactcat	tttaacaagg	Lagectigate	2736 2796
togeteeget carttgaatg	ttaggtttca	tagaactata	Ctaatcttct	Cacaaaayyt	
chahaaaata caqtcqttqa a	aaaaaatttt	gtatcaaaat	gtttggaaaa	ttagaagett	2856 2916
atcettaace totattoata	ctgacttgaa	ttattttcta	aaattaagag	Coglatacce	
acceptable cttttcacat	atcatttaaa	cttttattta	tattattact	gatttacage	2976 3036
ttagttatta attiticiti	ataagaatgc	catcaatata	catgetttta	tgtttttag	3036
aaaagggtgt gtttggatga	aagtaaaaaa	aaaaataaaa	tettteactg	loccitaatgg	3156
atatactatt taacattttt	tgaccctaaa	attcaccaac	agtctcccag	Lacatadac	3216
aggettaatg actggccctg	cattcttcac	aatattttc	cctaagcttt	gagcaaagtt	
ttasaaaaat acactaaaat	aatcaaaact	gttaagcagt	atattagtti	gyttatataa	3276
atteatetee aatttataae	atocatoocc	gatgttaatt	tgcttggcaa	Licigiaali	3336
attaagtgat ctcagtgaaa	catotcaaat	gccttaaatt	. aactaagttg	gtgaataaaa	3396
ataggatet agetaactet	tacaccatac	atactgatag	ttttcatat	gillicatill	3456 3516
cotatasttt ttaaaattta	gagtggcaac	aattttgctt	. aatatgggtt	acataagett	
tattttttcc tttgttcata	attatattct	. ttgaataggt	ctgtgtcaal	. Caagigaici	3576
anctagactg atcatagata	gaaggaaata	aggccaagtt	: caagaccagc	ctyggcaaca	3636
tatogagae ctototacaa	aaaaattaaa	l aaaaattago	caggcatgg:	ggcgtacact	3696
asatsattta teccaactac	t.cagaaaaat	gaggtgggag	gategetica	i gcccaggagg	3756
ttgagattgc agtgagccat	ggacatacca	a ctgcactaca	a gcctaggtaa	a cagcacgaga	3816
<del>-</del> -					

```
ccccaactct tagaaaatga aaaggaaata tagaaatata aaatttgctt attatagaca
                                                                   3876
cacagtaact cccagatatg taccacaaaa aatgtgaaaa gagagagaaa tgtctaccaa
agcagtattt tgtgtgtata attgcaagcg catagtaaaa taattttaac cttaatttgt
                                                                   3996
tittagtagt gittagattg aagattgagt gaaatatttt ettggcagat atteegtate
                                                                   4056
tggtggaaag ctacaatgca atgtcgttgt agttttgcat ggcttgcttt ataaacaaga
                                                                   4116
ttttttctcc ctccttttgg gccagttttc attacgagta actcacactt tttgattaaa
                                                                   4176
gaacttgaaa ttacgttatc acttagtata attgacatta tatagagact atgtaacatg
                                                                   4236
caatcattag aatcaaaatt agtactttgg tcaaaatatt tacaacattc acatacttgt
                                                                   4296
caaatattca tgtaattaac tgaatttaaa accttcaact attatgaagt gctcgtctgt
                                                                   4356
acaatcgcta atttactcag tttagagtag ctacaactct tcgatactat catcaatatt
                                                                   4476
tgacatcttt tccaatttgt gtatgaaaag taaatctatt cctgtagcaa ctggggagtc
atatatgagg tcaaagacat ataccttgtt attataatat gtatactata ataatagctg
                                                                   4536
gttatcctga gcaggggaaa aggttatttt taggaaaacc acttcaaata gaaagctgaa
                                                                   4596
gtacttctaa tatactgagg gaagtataat atgtggaaca aactctcaac aaaatgttta
                                                                    4656
ttgatgttga tgaaacagat cagtttttcc atccggatta ttattggttc atgattttat
                                                                   4716
atgtgaatat gtaagatatg ttctgcaatt ttataaatgt tcatgtcttt ttttaaaaaa
                                                                   4776
ggtgctattg aaattctgtg tctccagcag gcaagaatac ttgactaact ctttttgtct
                                                                    4836
ctttatggta ttttcagaat aaagtctgac ttgtgttttt gagattattg gtgcctcatt
                                                                    4896
                                                                    4932
aattcagcaa taaaggaaaa tatgcatctc aaaaat
<210> 117
<211> 4682
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<222> 31..33
<223> ATG
<221> misc_feature
<222> 301..303
<223> TGA
<221> polyA_signal
<222> 4655..4660
<223> AATAAA
 <400> 117
etgetgtece tggtgeteca caegtactee atg ege tae etg etg eec age gte
                                                                      54
                                 Met Arg Tyr Leu Leu Pro Ser Val
 gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg
                                                                     102
 Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
                        15
 egg etg etc tee gee tte etg ecc gee ege tte tae caa geg etg gae
                                                                     150
 Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
                     30
 25
 gac egg etg tae tge gte tae eag age atg gtg etc tte tte tte gag
                                                                     198
 Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Glu
                                    50
 aat tac acc ggg gtc cag aat ttc tct gca aag aat gtc caa aaa ttc
                                                                     246
 Asn Tyr Thr Gly Val Gln Asn Phe Ser Ala Lys Asn Val Gln Lys Phe
             60
                                 65
                                                                      294
 ata ttc aca ttg atc gta tcg aca aaa aag atg tcc cag aag aac aag
 Ile Phe Thr Leu Ile Val Ser Thr Lys Lys Met Ser Gln Lys Asn Lys
                             80
                                                                      343
 Asn Ile
 gcttatagaa ttttatgagt caccagatcc agaaagaaga aaaagatttc ctgggaaaag
                                                                      403
 tgttaattcc aaattaagta tcaagaagac tttaccatca atgttgatct taagtggttt
                                                                      463
 gactgcaggc atgcttatga ccgatgctgg aaggaagctg tatgtgaaca cctggatata
                                                                      523
 tggaacccta cttggctgcc tgtgggttac tattaaagca tagacaagta gctgtctcca
                                                                      583
  gacagtggga tgtgctacat tgtctatttt tggcggctgc acatgacatc aaattgtttc
                                                                      643
  ctgaatttat taaggagtgt aaataaagcc ttgttgattg aagattggat aatagaattt
                                                                      703
```

gtgacgaaag	ctgatatgca	atggtcttgg	gcaaacatac	ctggttgtac	aactttagca	763
togggggtgc	tagaagggta .	aaagctaaat	ggagtttctc	ctgctctgtc	Callectal	823
~aactaatra	caacttgaga .	addctdddad -	gattgtgtat	tttgcaagtc	agatggttgt	883
atttttaacc	attaatttgc	acctatttc	actttttctg	ttatttttaa	LLLactacaa	943
cttgacagct	ccaagetett	attactaaaq	tatttagtat	cttgcagcta	gilaalalii	1003
catcttttcc	tratttctac	aagtcagtga	aataaattgt	atttaggaag	tyttayyaty	1063
ttcaaaaaaa	agggtaaaaa	atattcataa	ggaaaaagct	ctgtttagca	Catgatttta	1123
ttatattaca	trattagetg	attttactca	ttttatattt	gcaaaataaa	tttctaatat	1183
ttattgaaat	toottaattt	gcacaccctg	tacacacaga	aaatggtata	aaatatgaga	1243
accaacttta	aaattgtgac	tctgattcat	tatagcagaa	Ctttaaattt	CCCagcitti	1303 1363
tgaagattta	agctacgcta	ttagtacttc	cctttgtctg	tgccataagt	gottgaaaac	1423
artaaaattt	totattttat	tttattttt	taatatcaaa	agagtcggtg	tgaacettyy	1483
ttggacccca	agttcacaag	atttttaagg	tgatgagagc	ctgcagacat	asttangaaa	1543
tttactagcg	tgtgcctttt	gcctgcttct	ctttgatttc	acagaatatt	gastagaag	1603
tcgcgtttct	gtagtgtggt	ggattcccac	tgggctctgg	tccttccctt	ggatecegte	1663
agtggtgctg	ctcagcggct	tgcacgtaga	cttgctagga	agaaatgcag	agecageceg	1723
tgctgcccac	tttcagagtt	gaactcttta	agcccttgtg	agtgggcttc	accayctact	1783
gcagaggcat	tttgcatttg	tctgtgtcaa	gaagttcacc	anttagge	agtgaaacta	1843
agacttaatt	cgtcatgact	gaacgaattt	gtttatttcc	cattaggttt	tcaccatttt	1903
cacattaata	tgtatcgcct	tagagcaaga	getgtgttee	ayyaaccaya	rrctattttt	1963
tagccatgga	acaatatatc	ccatgggaga	agacctttca	graceattr	ctacttacta	2023
gtgttataat	ttaaacttcg	attteeteat	agiccittaa	ttaaacatac	cagttggatc	2083
ctactggatt	tttgctgcag	aaatatatca	graggeecaca	taagtgactg	tattacacta	2143
atgataagca	aaatgaaaga	aacaacyacc	aayyyaaaac	tttgaagagt	catataatat	2203
cttctcccat	gccagagaat	addetettee	adjusticati	ggacactcag	gatatagttg	2263
gaattggttt	gtgtacatta cggggcatgg	gaatgtatge	trasastttc	ctcatgctga	attotaattt	2323
gcctaataat	gtaaagtaaa	gtaaaactta	attccatgtc	rrrattaagt	acagggattt	2383
tctcttacct	gtaaagtaaa gaatataatg	actilagatea	aaatttgaac	tttgagaggc	aatactgttg	2443
aatatattt	gaatataatg gattctaact	cattttaaca	aggtagcctg	acctgcataa	gatcacttga	2503
gaattatgtg	tcatagaact	atactaatct	tctcacaaaa	ggtctataaa	atacagtcgt	2563
acgccaggcc	tttgtatcaa	aatgtttgga	aaattagaag	cttctcctta	acctgtattg	2623
atactoactt	· waattatttt	ctaaaattaa	gageegtata	cctacctgta	agicilica	2683
catatoattt	· aaacttttgt	ttotattatt	actgatttac	agettagtta	Claacecee	2743
tttataagaa	taccatcaat	gtgcatgctt	ttatgtttt	cagaaaayyy	tgtgtttgga	2803
trasartasa	a aaaaaaaata	aaatctttca	ctgtctctaa	tggctgtgct	gillaacacc	2863
++++daccct	- aaaattcacc	aacagtetee	cagtacataa	aataggctta	acgaccygec	2923
cracattett	- cacaatattt	ttccctaaqc	tttgagcaaa	gttttaaaa	adiacactaa	2983
aataatcaaa	actottaago	agtatattag	tttggttata	taaattcatc	igcaatttat	3043
aagatgcatg	r occoatotta	atttacttaa	caattctgta	atcattaay	. gatettagtg	3103
aaacatotoa	a aatgccttaa	attaactaaq	ttggtgaata	aaagtgccga	cciggotaac	3163
tottacacca	- tacatactga	tagttttca	tatgtttcat	ttccatgiga	Littaaaac	3223 3283
ttagagtgg	- aacaattttd	r cttaatatqq	gttacataag	f ctttattti	Localigate	3343
ataattatai	r totttgaata	aatctatata	: aatcaagtga	l totaactaya	Cigateatay	3403
atamaamma	a ataaddccaa	gttcaagacg	: agcctgggca	i acatalogas	aaccigicia	3463
caaaaaaat	t aaaaaaaatt	. agccaggcat	ggtggcgtac	actgagtagt	ttgtcccagc	3523
tactcggga	g ggtgaggtgg	gaggateget	: tcagcccagg	aggityaya	tgcagtgagc	3583
catggacat	a ccactgcact	: acagcctagg	taacagcac	agaccccaa	tcttagaaaa	3643
tgaaaagga	a atatagaaat	ataaaatttg	, citattatas	, caaaccagt	a actcccagat	3703
atgtaccac	a aaaaatgtga	aaagagagag	, addigitizat	tadagtagtagtagtagtagtagtagtagtagtagtagtagta	ttttgtgtgt	3763
ataattgca	a gegeatagta	a aaataatti	aaccccaac	t atctddtag	t agtgtttaga a aagctacaat	3823
ttgaagatt	g agtgaaatat	- astanatta	tttataaac	a agattttt	c tccctccttt	3883
gcaatgtcg	t tgragttit	g catggetty	ctttttatt	r aaagaactt	g aaattacgtt	3943
tgggccagt	t ttcattacg	a gradereace	r actatotaa	r atgcaatca	t tagaatcaaa	4003
atcacttag	t ataattgac	z ttatataya; F stttscssc:	a ticacatac	t totcaaata	t tcatgtaatt	4063
attagtact	t pagedada	a actattato	a agtgctcgt	c totacaatc	g ctaatttact	4123
aactgaatt	a tagetacaa	c tettemata	c tatcatcaa	t atttgacat	c ttttccaatt	4183
	a sagtasato	t attectota:	a caactaaaa	a gtcatatat	g aggicaaaya	4243
catatacct	+ ~++a++a+a	a tatotatac	t ataataata	g ctggttatc	c cgagcayyyy	4303
aaaaaatta	at ttttaggaa	a accacttca	a atagaaagc	t gaagtactt	c taatatactg	4363
aaaayyee		•	= -			

```
agggaagtat aatatgtgga acaaactctc aacaaatgt ttattgatgt tgatgaaaca
                                                                 4423
                                                                 4483
gatcagtttt tccatccgga ttattattgg ttcatgattt tatatgtgaa tatgtaagat
atgttctgca attttataaa tgttcatgtc tttttttaaa aaaggtgcta ttgaaattct
                                                                 4543
gtgtctccag caggcaagaa tacttgacta actctttttg tctctttatg gtattttcag
                                                                 4603
aataaagtet gaettgtgtt tttgagatta ttggtgeete attaatteag caataaagga
                                                                 4663
                                                                  4682
aaatatgcat ctcaaaaat
<210> 118
<211> 4558
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<222> 31..33
<223> ATG
<221> misc_feature
<222> 235..237
<223> TGA
<221> polyA_signal
<222> 4531..4536
<223> AATAAA
<400> 118
etgetgtece tggtgeteca caegtactee atg ege tae etg etg ece age gte
                                                                    54
                                Met Arg Tyr Leu Leu Pro Ser Val
                                1
gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg
                                                                   102
Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
                       15
egg etg etc tee gee tte etg eee gee ege tte tae caa geg etg gae
                                                                   150
Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
                                       35
 25
gac egg etg tae tge gte tae cag age atg gtg etc tte tte tte gag
                                                                   198
Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Glu
                                   50
                45
aat tac acc ggg gtc cag gat gct tat aga att tta tga gtcaccagat
                                                                   247
Asn Tyr Thr Gly Val Gln Asp Ala Tyr Arg Ile Leu
            60
 ccagaaagaa gaaaaagatt tcctgggaaa agtgttaatt ccaaattaag tatcaagaag
                                                                   307
 actitaccat caatgitgat citaagiggt tigacigcag gcatgcitat gaccgatgci
                                                                   367
 ggaaggaagc tgtatgtgaa cacctggata tatggaaccc tacttggctg cctgtgggtt
                                                                   427
                                                                   487
 actattaaag catagacaag tagctgtctc cagacagtgg gatgtgctac attgtctatt
 ttiggegget geacatgaea teaaattgtt teetgaattt attaaggagt gtaaataaag
                                                                   547
 ccttgttgat tgaagattgg ataatagaat ttgtgacgaa agctgatatg caatggtctt
                                                                   607
 gggcaaacat acctggttgt acaactttag catcggggct gctggaaggg taaaagctaa
                                                                   667
 atggagtttc tcctgctctg tccatttcct atgaactaat gacaacttga gaaggctggg
                                                                   727
 aggattgtgt attttgcaag tcagatggct gcatttttga gcattaattt gcagcgtatt
                                                                   787
 tcactttttc tgttattttc aatttattac aacttgacag ctccaagctc ttattactaa
                                                                    847
 agtatttagt atottgcagc tagttaatat ttcatctttt gcttatttct acaagtcagt
                                                                   907
 gaaataaatt gtatttagga agtgtcagga tgttcaaagg aaagggtaaa aagtgttcat
                                                                   967
 1027
 cattttatat ttgcaaaata aatttctaat atttattgaa attgcttaat ttgcacaccc
                                                                   1087
 tgtacacaca gaaaatggta taaaatatga gaacgaagtt taaaattgtg actctgattc
                                                                   1147
 attatagcag aactttaaat ttcccagctt tttgaagatt taagctacgc tattagtact
                                                                   1207
 1267
 tttaatatca aaagagtcgg tgtgaacctt ggttggaccc caagttcaca agatttttaa
                                                                   1327
 ggtgatgaga gcctgcagac attctgccta gatttactag cgtgtgcctt ttgcctgctt
 ctctttgatt tcacagaata ttcattcaga agtcgcgttt ctgtagtgtg gtggattccc
 actgggctct ggtccttccc ttggatcccg tcagtggtgc tgctcagcgg cttgcacgta
                                                                   1507
  gacttgctag gaagaaatgc agagccagcc tgtgctgccc actttcagag ttgaactctt
                                                                   1567
  taagecettg tgagtggget teaceageta etgeagagge attttgeatt tgtetgtgte
                                                                   1627
  aagaagttca ccttctcaag ccagtgaaat acagacttaa ttcgtcatga ctgaacgaat
                                                                   1687
  ttgtttattt cccattaggt ttagtggagc tacacattaa tatgtatcgc cttagagcaa
                                                                   1747
```

```
gagetgtgtt ccaggaacca gatcacgatt tttagecatg gaacaatata teccatggga
                                                                    1807
gaagaccttt cagtgtgaac tgttctattt ttgtgttata atttaaactt cgatttcctc
                                                                    1867
atagteettt aagttgacat ttetgettae tgetaetgga tttttgetge agaaatatat
                                                                    1927
cagtggccca cattaaacat accagttgga tcatgataag caaaatgaaa gaaataatga
                                                                    1987
ttaagggaaa attaagtgac tgtgttacac tgcttctccc atgccagaga ataaactctt
                                                                    2047
tcaagcatca tctttgaaga gtcgtgtggt gtgaattggt ttgtgtacat tagaatgtat
                                                                    2107
qcacacatcc atggacactc aggatatagt tggcctaata atcggggcat gggtaaaact
                                                                    2167
tatgaaaatt tcctcatgct gaattgtaat tttctcttac ctgtaaagta aaatttagat
                                                                    2227
caattccatg totttgttaa gtacagggat ttaatatatt ttgaatataa tgggtatgtt
                                                                    2287
                                                                    2347
ctaaatttga actttgagag gcaatactgt tggaattatg tggattctaa ctcattttaa
caaggtagcc tgacctgcat aagatcactt gaatgttagg tttcatagaa ctatactaat
                                                                    2407
cttctcacaa aaggtctata aaatacagtc gttgaaaaaa attttgtatc aaaatgtttg
                                                                     2467
gaaaattaga agcttctcct taacctgtat tgatactgac ttgaattatt ttctaaaatt
                                                                     2527
aagagccgta tacctacctg taagtctttt cacatatcat ttaaactttt gtttgtatta
                                                                    2587
                                                                     2647
ttactgattt acagcttagt tattaatttt tctttataag aatgccgtcg atgtgcatgc
ttttatgttt ttcagaaaag ggtgtgtttg gatgaaagta aaaaaaaaa taaaatcttt
                                                                     2707
cactgtctct aatggctgtg ctgtttaaca ttttttgacc ctaaaaattca ccaacagtct
                                                                     2767
                                                                     2827
cccagtacat aaaataggct taatgactgg ccctgcattc ttcacaatat ttttccctaa
                                                                     2887
gctttgagca aagttttaaa aaaatacact aaaataatca aaactgttaa gcagtatatt
agtttggtta tataaattca tctgcaattt ataagatgca tggccgatgt taatttgctt
                                                                     2947
                                                                     3007
ggcaattctg taatcattaa gtgatctcag tgaaacatgt caaatgcctt aaattaacta
agttggtgaa taaaagtgcc gatctggcta actcttacac catacatact gatagttttt
                                                                     3067
catatgtttc atttccatgt gatttttaaa atttagagtg gcaacaattt tgcttaatat
                                                                    3127
qqqttacata aqctttattt tttcctttgt tcataattat attctttgaa taggtctgtg
                                                                     3187
tcaatcaagt gatctaacta gactgatcat agatagaagg aaataaggcc aagttcaaga
                                                                     3247
ccagcctggg caacatatcg agaacctgtc tacaaaaaaa ttaaaaaaa ttagccaggc
                                                                     3307
                                                                     3367
atggtggcgt acactgagta gtttgtccca gctactcggg agggtgaggt gggaggatcg
cttcagccca ggaggttgag attgcagtga gccatggaca taccactgca ctacagccta
                                                                     3427
ggtaacagca cgagacccca actcttagaa aatgaaaagg aaatatagaa atataaaatt
                                                                     3487
tgcttattat agacacacag taactcccag atatgtacca caaaaaatgt gaaaagagag
                                                                     3547
agaaatgtct accaaagcag tattttgtgt gtataattgc aagcgcatag taaaataatt
                                                                     3607
ttaaccttaa tttgttttta gtagtgttta gattgaagat tgagtgaaat attttcttgg
                                                                     3667
cagatattcc gtatctggtg gaaagctaca atgcaatgtc gttgtagttt tgcatggctt
                                                                     3727
gctttataaa caagattttt tctccctcct tttgggccag ttttcattac gagtaactca
                                                                     3787
                                                                     3847
cactttttga ttaaagaact tgaaattacg ttatcactta gtataattga cattatatag
                                                                     3907
agactatgta acatgcaatc attagaatca aaattagtac tttggtcaaa atatttacaa
cattcacata cttgtcaaat attcatgtaa ttaactgaat ttaaaacctt caactattat
                                                                     3967
gaagtgctcg tctgtacaat cgctaattta ctcagtttag agtagctaca actcttcgat
                                                                     4027
actatcatca atatttgaca tcttttccaa tttgtgtatg aaaagtaaat ctattcctgt
                                                                     4087
agcaactggg gagtcatata tgaggtcaaa gacatatacc ttgttattat aatatgtata
                                                                     4147
                                                                     4207
ctataataat agctggttat cctgagcagg ggaaaaggtt atttttagga aaaccacttc
aaatagaaag ctgaagtact tctaatatac tgagggaagt ataatatgtg gaacaaactc
                                                                     4267
                                                                     4327
tcaacaaaat gtttattgat gttgatgaaa cagatcagtt tttccatccg gattattatt
ggttcatgat tttatatgtg aatatgtaag atatgttctg caattttata aatgttcatg
                                                                     4387
totttttta aaaaaggtgo tattgaaatt otgtgtotoo agcaggoaag aatacttgao
                                                                     4447
taactctttt tgtctcttta tggtattttc agaataaagt ctgacttgtg tttttgagat
                                                                     4507
                                                                     4558
tattggtgcc tcattaattc agcaataaag gaaaatatgc atctcaaaaa t
<210> 119
<211> 5270
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<222> 31..33
 <223> ATG
 <221> misc_feature
 <222> 229..231
 <223> TAG
 <221> polyA_signal
 <222> 5243..5248
 <223> AATAAA
```

<pre>&lt;400&gt; 119 ctgctgtccc tggtgctcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc</pre>	54
Met Arg Tyr Leu Leu Pro Ser Val	
1 5	
gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg	102
gtg ctc ctg ggg acg gcg ccc acc tac gtg ttg gcc tag gas gas as	
Val Leu Cly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp	
10	150
egg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac	130
Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp	
75 1 30	198
gac egg etg tac tge gte tac eag age atg gtg ete tte tte tgag	190
Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Glu	
, <b>4</b>	251
aat tac acc ggg gtc cag aga ttg gat tcg tag attaaacttg agaaacaaac	271
Asn Tyr Thr Gly Val Gln Arg Leu Asp Ser *	
60 65	311
cataaaagtg gaaggccctc tttaacaata ttgctatatg gagatttgcc aaaaaataaa	371
gaaaatataa tatatttagc aaatcatcaa agcacagttg actggattgt tgctgacatc	431
ttggccatca ggcagaatgc gctaggacat gtgcgctacg tgctgaaaga agggttaaaa	
tggctgccat tgtatgggtg ttactttgct cagcatggag gaatctatgt aaagcgcagt	491
gccaaattta acgagaaaga gatgcgaaac aagttgcaga gctacgtgga cgcaggaact	551
ccaatgtatc ttgtgatttt tccaqaaggt acaaggtata atccagagca aacaaaagtc	611
ctttcagcta gtcaggcatt tgctgcccaa cgtggccttg cagtattaaa acatgtgcta	671
acaccacgaa taaaggcaac tcacgttgct tttgattgca tgaagaatta tttagatgca	731
atttatgatg ttacqqtqqt ttatgaaggg aaagacgatg gagggcagcg aagagagtca	791
ccgaccatga cggaatttct ctgcaaagaa tgtccaaaaa ttcatattca cattgatcgt	851
atcgacaaaa aagatgtccc agaagaacaa gaacatatga gaagatggct gcatgaacgt	911
ttcgaaatca aagataagat gcttatagaa ttttatgagt caccagatcc agaaagaaga	971
aaaagatttc ctgggaaaag tgttaattcc aaattaagta tcaagaagac tttaccatca	1031
atgttgatct taagtggttt gactgcaggc atgcttatga ccgatgctgg aaggaagctg	1091
tatgtgaaca cotggatata tggaaccota ottggotgoo tgtgggttac tattaaagca	1151
tagacaagta getgteteea gacagtggga tgtgetacat tgtetatttt tggeggetge	1211
acatgacatc aaattgtttc ctgaatttat taaggagtgt aaataaagcc ttgttgattg	1271
acatgacatc adattyttte etgaattat taaggagtge aastatge aagattggat aatagaattt gtgacgaaag etgatatgca atggtettgg gcaaacatac	1331
aagattggat aatagaatt gtgddgdadg tagaaggta aaggtaaat ggagttgt	1391
ctggttgtac aactttagca tcggggctgc tggaagggta aaagctaaat ggagtttctc	1451
ctgctctgtc catttcctat gaactaatga caacttgaga aggctgggag gattgtgtat	1511
tttgcaagtc agatggctgc atttttgagc attaatttgc agcgtatttc actttttctg	1571
ttatttcaa tttattacaa cttgacagct ccaagctctt attactaaag tatttagtat	1631
cttgcagcta gttaatattt catcttttgc ttatttctac aagtcagtga aataaattgt	1691
atttaggaag tgtcaggatg ttcaaaggaa agggtaaaaa gtgttcatgg ggaaaaagct	1751
ctgtttagca catgatttta ttgtattgcg ttattagctg attttactca ttttatattt	1811
gcaaaataaa tttctaatat ttattgaaat tgcttaattt gcacaccctg tacacacaga	1871
aaatggtata aaatatgaga acgaagttta aaattgtgac tctgattcat tatagcagaa	1931
ctttaaattt cccagctttt tgaagattta agctacgcta ttagtacttc cctttgtctg	1991
tgccataagt gcttgaaaac gttaaggttt tctgttttgt tttgttttt taatatcaaa	2051
arantennth thaacetton ttonacecca anticacady accellady tyacyayay	2111
ctgcagacat tctgcctaga tttactagcg tgtgcctttt gcctgcttct ctttgatttc	2171
acagaatatt cattcagaag tcgcgtttct gtagtgtggt ggattcccac tgggctctgg	2231
teetteett ggateegte agtggtgetg eteagegget tgeaegtaga ettgetagga	2231
agaaatgcag agccagcctg tgctgcccac tttcagagtt gaactcttta agcccttgtg	2351
agtgggcttc accagctact gcagaggcat tttgcatttg tctgtgtcaa gaagttcacc	
trotcaagec agreeaatac agacttaatt cgtcatgact gaacgaattt gtttatttee	2411
cattaggett agtggageta cacattaata tgtategeet tagageaaga getgtgttee	2471
aggaaccaga tcacgatttt tagccatgga acaatatatc ccatgggaga agacctttca	2531
grangacta trotatitit grattataat tiaaacticg atticcicat agreciitaa	2591
gttgacattt ctrcttactg ctactggatt tttgctgcag aaatatatca gtggcccaca	2651
ttaaacatac cagttggatc atgataagca aaatgaaaga aataatgatt aagggadadt	2711
taagtgactg tottagactg cttctcccat gccagagaat aaactctttc aagcatcatc	2771
tttgaagagt cotground gaattggttt gtgtacatta gaatgtatgc acacatccat	2831
gracacteag gatatagttg gcctaataat cggggcatgg gtaaaactta tgaaaatttc	2891
ctcatgctga attgtaattt tctcttacct gtaaagtaaa	2951

```
tttgttaagt acagggattt aatatatttt gaatataatg ggtatgttct aaatttgaac
                                                                   3011
tttgagaggc aatactgttg gaattatgtg gattctaact cattttaaca aggtagcctg
                                                                   3071
acctgcataa gatcacttga atgttaggtt tcatagaact atactaatct tctcacaaaa
                                                                   3131
ggtctataaa atacagtcgt tgaaaaaaat tttgtatcaa aatgtttgga aaattagaag
                                                                   3191
cttctcctta acctgtattg atactgactt gaattatttt ctaaaattaa gagccgtata
                                                                   3251
cctacctgta agtcttttca catatcattt aaacttttgt ttgtattatt actgatttac
                                                                   3311
agcttagtta ttaatttttc tttataagaa tgccgtcgat gtgcatgctt ttatgttttt
                                                                   3371
3431
tggctgtgct gtttaacatt ttttgaccct aaaattcacc aacagtctcc cagtacataa
                                                                   3491
aataggetta atgactggee etgeattett cacaatattt tteeetaage tttgagcaaa
                                                                   3551
gttttaaaaa aatacactaa aataatcaaa actgttaagc agtatattag tttggttata
                                                                   3611
taaattcatc tgcaatttat aagatgcatg gccgatgtta atttgcttgg caattctgta
                                                                   3671
atcattaagt gatctcagtg aaacatgtca aatgccttaa attaactaag ttggtgaata
                                                                   3731
aaagtgccga tctggctaac tcttacacca tacatactga tagtttttca tatgtttcat
                                                                   3791
ttccatgtga tttttaaaat ttagagtggc aacaattttg cttaatatgg gttacataag
                                                                   3851
ctttattttt tcctttgttc ataattatat tctttgaata ggtctgtgtc aatcaagtga
                                                                   3911
tctaactaga ctgatcatag atagaaggaa ataaggccaa gttcaagacc agcctgggca
                                                                   3971
acatatcgag aacctgtcta caaaaaaatt aaaaaaaatt agccaggcat ggtggcgtac
                                                                    4031
actgagtagt ttgtcccagc tactcgggag ggtgaggtgg gaggatcgct tcagcccagg
                                                                    4091
aggttgagat tgcagtgagc catggacata ccactgcact acagcctagg taacagcacg
                                                                    4151
agaccccaac tettagaaaa tgaaaaggaa atatagaaat ataaaattig ettattatag
                                                                    4211
acacacagta acteccagat atgtaccaca aaaaatgtga aaagagagag aaatgtetac
                                                                    4271
caaagcagta ttttgtgtgt ataattgcaa gcgcatagta aaataatttt aaccttaatt
                                                                    4331
tgtttttagt agtgtttaga ttgaagattg agtgaaatat tttcttggca gatattccgt
                                                                    4391
arctggtgga aagctacaat gcaatgtcgt tgtagttttg catggcttgc tttataaaca
                                                                    4451
agattttttc tccctccttt tgggccagtt ttcattacga gtaactcaca ctttttgatt
                                                                    4511
aaagaacttg aaattacgtt atcacttagt ataattgaca ttatatagag actatgtaac
                                                                    4571
atgcaatcat tagaatcaaa attagtactt tggtcaaaat atttacaaca ttcacatact
                                                                    4631
tgtcaaatat tcatgtaatt aactgaattt aaaaccttca actattatga agtgctcgtc
                                                                    4691
tgtacaatcg ctaatttact cagtttagag tagctacaac tcttcgatac tatcatcaat
                                                                    4751
atttgacatc ttttccaatt tgtgtatgaa aagtaaatct attcctgtag caactgggga
                                                                    4811
gtcatatatg aggtcaaaga catatacctt gttattataa tatgtatact ataataatag
                                                                    4871
ctggttatcc tgagcagggg aaaaggttat ttttaggaaa accacttcaa atagaaagct
                                                                    4931
 gaagtacttc taatatactg agggaagtat aatatgtgga acaaactctc aacaaaatgt
                                                                    4991
 ttattgatgt tgatgaaaca gatcagtttt tccatccgga ttattattgg ttcatgattt
                                                                    5051
 tatatgtgaa tatgtaagat atgttctgca attttataaa tgttcatgtc tttttttaaa
                                                                    5111
 aaaggtgcta ttgaaattct gtgtctccag caggcaagaa tacttgacta actctttttg
                                                                    5171
 tototttatg gtattttcag aataaagtot gacttgtgtt tttgagatta ttggtgcctc
                                                                    5231
                                                                    5270
 attaattcag caataaagga aaatatgcat ctcaaaaat
 <210> 120
 <211> 5002
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <222> 31..33
 <223> ATG
 <221> misc_feature
 <222> 322..324
 <223> TAA
 <221> polyA_signal
 <222> 4975..4980
  <223> AATAAA
  <400> 120
                                                                       54
  ctgctgtccc tggtgctcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc
                                  Met Arg Tyr Leu Leu Pro Ser Val
                                  1
  gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg
                                                                      102
  Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
      10
  egg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac
                                                                      150
```

								_						•		
Arg	Leu	Leu	Ser	Ala		Leu	Pro	Ala	Arg	Phe	Tyr	Gin	Ala	Leu	Asp 40	
25					30					35				++0		198
gac	cgg	ctg	tac	tgc	gtc	tac	cag	agc	atg	gtg	TON	Dho	Dho	Dho	Glu	190
Asp	Arg	Leu	Tyr		Val	Tyr	GIN	Ser	мес 50	vai	ьец	FILE	FILE	55	Giu	
				45	~~~	ata	++~	cta		aga	gat	tta	сса		aat	246
aat	tac	acc	ggg	gtc	cag	Ile	LLG	Len	Tur	Glv	Asn	Leu	Pro	Lvs	Asn	
Asn	Tyr	Tnr		vai	GIII	116	пеп	65	ıyı	Gry	АЗР	<b></b>	70	27.5		
			60	a+ a	tat	tta	aca		cat	саа	agc	aca		gta	tct	294
aaa	gaa	Aar	TIO	Tla	Tur	Leu	Ala	Asn	His	Gln	Ser	Thr	Asp	Val	Ser	
гÃг	GIU	75	TIE	116	тУт	пец	80	11011		0		85				
+~+	~~+	/J	tac	ana	add	tac		σta	taa	tcca	agag	caa	acaa	aagt	cc	344
Cyc	Acn	Dhe	Ser	Ara	Ara	Tyr	Lvs	Val	*					_		
Cys	90	1110	DCI	*** 9	5	95	-1-									
+++	cade	tag	tcag	acat	tt a	ctgc	ccaa	c gt	ggcc	ttgc	agt	atta	aaa	catg	tgctaa	404
cac	cacd	aat .	aaag	acaa	ct c	acati	tacti	t tt	gatt	gcat	gaa	gaat	tat	ttag	atgcaa	464
+++	at da:	tat	tacq	ataa	tt t	atga	aggg	a aa	gacg	atgg	agg	gcag	cga	agag	agtcac	524
cga	ccat	gac .	agaa	tttc	tc t	gcaa	agaa	t gt	ccaa	aaat	tca	tatt	cac	attg	atcgta	584
tca	acaa.	940 888	agat	atcc	ca d	aaga	acaa	g aa	cata	tgag	aag	atgg	ctg	catg	aacgtt	644
tca	aaat	caa	agat	aada	ta c	ttat	agaa	t tt	tatg	agtc	acc	agat	cca	gaaa	gaagaa	704
222	catt	tcc	taˈaa	aaaa	at a	rttaa	ttcc	a aa	ttaa	gtat	caa	gaag	act	ttac	catcaa	764
tat	taat	ctt	aagt	aatt	ta a	ctac	aggc	a to	ctta	tgac	cga	tgct	gga	agga	agctgt	824
ato	taaa	cac	ctaa	atat	at o	rgaac	ccta	c tt	ggct	gcct	gtg	ggtt	act	atta	aagcat	884
ara	caad	tag	ctat	ataa	acra	cagt	aaaa	t at	gcta	catt	gtc	tatt	ttt	ggcg	gctgca	944
cat	gaca	tca	aatt	attt	cc t	gaat	ttat	t aa	ggag	tgta	aat	aaag	cct	tgtt	gattga	1004
aga	ttaa	ata	atag	aatt	ta t	gacq	aaaq	c to	ratat	gcaa	tgg	tctt	ggg	caaa	catacc	1064
taa	ttat	aca	actt	tago	at c	aaaa	ctac	t ac	aagg	gtaa	aag	ctaa	atg	gagt	ttctcc	1124
tac	teta	tcc	attt	ccta	ta a	acta	atga	c aa	ctta	agaa	ggc	tggg	agg	attg	tgtatt	1184
++0	caad	tca	acta	acta	ca t	tttt	gage	a tt	aatt	tgca	gcg	tatt	tca	cttt	ttctgt	1244
tat	+++~	aat	ttat	taca	ac t	tgac	acct	C Ca	agct	ctta	tta	ctaa	agt	attt	agtatc	1304
++0	carc	tar	ttaa	tatt	to a	atctt	ttac	t ta	attto	taca	agt	cagt	gaa	ataa	attgta	1364
+++	anna	aat	atca	ggat	at t	caaa	ggaa	a qo	gtaa	ıaaag	tgt	tcat	:ggg	gaaa	laagctc	1424
tat	ttad	cac	atga	tttt	at t	gtat	taca	t ta	ittag	rctga	ttt	tact	cat	ttta	tatttg	1484
caa	aata	aat	ttct	aata	itt t	tattq	aaat	t go	cttaa	itttg	cac	acco	tgt	acac	cacagaa	1544
aat	anta	taa	aata	tgac	raa o	caaaa	ttta	a aa	attgt	gact	ctg	ratto	catt	atag	gcagaac	1604
+++	aaat	ttc	ccac	rcttt	tt d	raaqa	ttta	a go	ctacg	gctat	tag	įtact	tcc	cttt	gtctgt	1664
acc	atas	ata	ctto	raaaa	aca 1	ttaao	rattt	t ci	tgttt	tgtt	ttg	gtttt	ttt	aata	atcaaaa	1724
gad	ttcac	rtat	gaag	ctto	rat 1	taaac	ccca	ia qi	ttcac	caaga	. ttt	ttaa	aggt	gate	gagagcc	1784
tac	agac	att	ctac	ctac	rat 1	ttact	agco	rt g	tgcct	itttg	cct	gctt	ctc	tttg	gatttca	1844
cac	raats	atto	atto	agaa	aat (	cacat	ttct	a ta	agtgt	iggtg	gat	tccc	cact	ggg	ctctggt	1904
cct	tccc	rtta	gato	accat	ca (	ataat	acto	ic to	cage	ggctt	gca	acgta	agac	ttg	ctaggaa	1964
αaa	ator	รอดล	acca	acct	at (	actac	ccac	t t	tcaga	agttg	aac	ctcti	taa	gcc	cttgtga	2024
ato	raact	tca	CCAC	actac	cta :	cagac	racat	t t	tgcai	tttgt	ctg	gtgto	caag	aag	ttcacct	2084
tot	caac	rcca	ata	aata	aca ·	gactt	aatt	c g	tcate	gacto	aa	cgaar	tttg	ttt	atttccc	2144
att	raddi	ttta	ata	rage	tac	acatt	aata	at g	tato	gcctt	: aga	agca	agag	ctg	tgttcca	2204
aa	aacca	dat	cabo	ratti	ttt	agcca	ataga	aa c	aatai	tatco	ca:	cggg	agaa	gac	ctttcag	2264
t or	taaa	ctat	tcta	attt	tta	tatta	ataat	tt t	aaac	ttcga	a tt	tact	cata	gtc	ctttaag	2324
++	raca'	tttc	tac	ttac	tac	tacto	rgati	tt t	tgct	gcaga	a aa	tata	tcag	tgg	cccacat	2384
t a	aaca	tacc	agt	taga	tca	tgata	aagca	aa a	atga	aagaa	a at	aatg	atta	agg	gaaaatt	2444
22	ataa	ctat	att:	acac	tac	ttctc	ccai	ta c	caga	gaata	a aa	ctct	ttca	agc	atcatct	2504
++	naan	agtc	ατά	taat	ata	aatt	aatt	ta t	gtac	attag	g aa	tgta	tgca	cac	atccatg	2564
αa	cact	cand	ata	tagt	taa	ccta	ataa	tc q	gggc	atggg	g ta	aaac	ttat	gaa	aatttcc	2624
tc	atoc	taaa	tta	taat	ttt	ctct	tacc	ta t	aaag	taaaa	a tt	taga	tcaa	ttc	catgtct	2684
t t	atta	anta	cad	aga t	tta	atat	attt	ta a	atat	aatgg	g gt	atgt	tcta	aat	ttgaact	2744
++	gaga	aaca	ata	ctat	taa	aatt	atqt	gg a	ittct	aacto	c at	ttta	acaa	ggt	agcctga	2804
	taca	taad	atc	actt	gaa	tatt	adat	tt c	ataq	aact	a ta	ctaa	tctt	ctc	acaaaag	2864
at	ctat	2222	tac	agto	att	gaaa	aaaa	tt t	:tata	tcaa	a at	gttt	ggaa	ı aat	tagaagc	2924
++	ctcc	ttaa	cct	atat	tga	tact	gact	ta a	atta	tttt	c ta	aaat	taag	ago	cgtatac	2984
ct	acct	ntaa	atc	++++	cac	atat	catt	ta a	actt	ttqt	t tg	tatt	atta	ı ctg	_l atttaca	3044
gc	ttag	ttat	: taa	tttt	tct	ttat	aaga	at o	gccgt	cgat	g tg	cate	cttt	. cat	gttttc	3104
ag	gaaaa	ıgggt	gtg	rtttg	gat	gaaa	gtaa	aa a	aaaaa	aata	a aa	icctt	ccac	: tgt	ctctaat	3164

```
ggctgtgctg tttaacattt tttgacccta aaattcacca acagtctccc agtacataaa
                                                                    3224
ataggettaa tgactggeee tgcattette acaatatttt teeetaaget ttgageaaag
                                                                    3284
ttttaaaaaa atacactaaa ataatcaaaa ctgttaagca gtatattagt ttggttatat
                                                                    3344
aaattcatct gcaatttata agatgcatgg ccgatgttaa tttgcttggc aattctgtaa
                                                                    3404
tcattaagtg atctcagtga aacatgtcaa atgccttaaa ttaactaagt tggtgaataa
aagtgccgat ctggctaact cttacaccat acatactgat agtttttcat atgtttcatt
                                                                    3524
tccatgtgat ttttaaaatt tagagtggca acaattttgc ttaatatggg ttacataarc
                                                                    3584
                                                                    3644
tttatttttt cctttgttca taattatatt ctttgaatag gtctgtgtca atcaagtgat
ctaactagac tgatcataga tagaaggaaa taaggccaag ttcaagacca gcctgggcaa
catatcgaga acctgtctac aaaaaaatta aaaaaaatta gccaggcatg gtggcgtaca
                                                                     3764
ctgagtagtt tgtcccagct actcgggagg gtgaggtggg aggatcgctt cagcccagga
                                                                    3824
ggttgagatt gcagtgagcc atggacatac cactgcacta cagcctaggt aacagcacga
                                                                    3884
gaccccaact cttagaaaat gaaaaggaaa tatagaaata taaaatttgc ttattataga
cacacagtaa ctcccagata tgtaccacaa aaaatgtgaa aagagagaga aatgtctacc
                                                                     4004
aaagcagtat tttgtgtgta taattgcaag cgcatagtaa aataatttta accttaattt
                                                                     4064
gtttttagta gtgtttagat tgaagattga gtgaaatatt ttcttggcag atattccgta
                                                                     4124
tctggtggaa agctacaatg caatgtcgtt gtagttttgc atggcttgct ttataaacaa
                                                                     4184
gattttttct ccctcctttt gggccagttt tcattacgag taactcacac tttttgatta
                                                                     4244
aagaacttga aattacgtta tcacttagta taattgacat tatatagaga ctatgtaaca
                                                                     4304
tgcaatcatt agaatcaaaa ttagtacttt ggtcaaaata tttacaacat tcacatactt
                                                                     4364
gtcaaatatt catgtaatta actgaattta aaaccttcaa ctattatgaa gtgctcgtct
gtacaatcgc taatttactc agtttagagt agctacaact cttcgatact atcatcaata
                                                                     4484
tttgacatct tttccaattt gtgtatgaaa agtaaatcta ttcctgtagc aactggggag
                                                                     4544
tcatatatga ggtcaaagac atataccttg ttattataat atgtatacta taataatagc
                                                                     4604
tggttatcct gagcagggga aaaggttatt tttaggaaaa ccacttcaaa tagaaagctg
                                                                     4664
aagtacttct aatatactga gggaagtata atatgtggaa caaactctca acaaaatgtt
                                                                     4724
tartgatgtt gatgaaacag atcagttttt ccatccggat tattattggt tcatgatttt
                                                                     4784
                                                                     4844
atatgtgaat atgtaagata tgttctgcaa ttttataaat gttcatgtct ttttttaaaa
aaggtgctat tgaaattctg tgtctccagc aggcaagaat acttgactaa ctctttttgt
ctctttatgg tattttcaga ataaagtctg acttgtgttt ttgagattat tggtgcctca
                                                                     4964
                                                                     5002
ttaattcagc aataaaggaa aatatgcatc tcaaaaat
<210> 121
<211> 4958
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<222> 31..33
<223> ATG
<221> misc_feature
<222> 577..579
<223> TGA
<221> polyA_signal
<222> 4931..4936
 <223> AATAAA
 <400> 121
 ctgctgtccc tggtgctcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc
                                                                       54
                                  Met Arg Tyr Leu Leu Pro Ser Val
                                  1
 gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg
                                                                      102
 Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
                         15
     10
 egg etg etc tee gee tte etg eee gee ege tte tae eaa geg etg gae
                                                                      150
 Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
                                         35
                     30
 gac egg etg tae tge gte tae eag age atg gtg ete tte tte tte gag
                                                                       198
 Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Glu
                 45
 aat tac acc ggg gtc cag ata ttg cta tat gga gat ttg cca aaa aat
                                                                       246
 Asn Tyr Thr Gly Val Gln Ile Leu Leu Tyr Gly Asp Leu Pro Lys Asn
                                 65
             60
```

Lys	Glu	Asn 75	Ile	Ile	Tyr	Leu	Ala 80	Asn	His	Gln	Ser	Thr 85	gtt Val	Asp	Trp	294
Ile	Val	Ala	Asp	Ile	Leu	Ala 95	Ile	Arg	Gln	Asn	100	Leu	gga Gly	HIS	val	342
Arg	tac Tyr	Val	Leu	Lys	Glu 110	Gly	Leu	Lys	Trp	Leu 115	Pro	Leu	tat Tyr	GIĀ	120	390
tac Tyr	Phe	Ala	Gln	His	Gly	Gly	Ile	Tyr	Val 130	Lys	Arg	Ser	gcc Ala	Lys 135	Pne	438
Asn	Glu	Lys	Glu 140	Met	Arg	Asn	Lys	Leu 145	Gln	Ser	Tyr	Val	gac Asp 150	Ala	GIĀ	486
act Thr	cca Pro	aat Asn 155	ttc	tct Ser	gca Ala	aag Lys	aat Asn 160	gtc Val	caa Gln	aaa Lys	ttc Phe	ata Ile 165	ttc Phe	aca Thr	ttg Leu	534
Ile	Val	tcg Ser	Thr	Lys	Lys	Met 175	Ser	Gln	Lys	Asn	Lys 180	Asn	ata Ile	*		579
gaa	aata	gct	gcat	gaac	gt t	tcga	aatc	a aa	gata	agat	gct	tata	gaa	tttt	atgagt	639
CAC	cada	tcc	aσaa	aσaa	ga a	aaag	attt	c ct	ggga	aaag	tgt	taat	tcc	aaat	taagta	699
tca	agaa	gac	ttta	ccat	ca a	tgtt	gatc	t ta	agtg	gttt	gac	tgca	ggc	atgc	ttatga	759
ccg	atgc	tgg	aagg	aagc	tg t	atgt	gaac	a cc	tgga	tata	tgg	aacc	cta	cttg	gctgcc	819 879
tgt	gggt	tac	tatt	aaag	ca t	agac	aagt	a gc	tgtc	tcca	gac	agtg	gga	tgtg	ctacat	939
tgt	ctat	ttt	tggc	ggct	gc a	catg	acat	c aa	attg	TTTC	ctg	aacc	Lat	ctas	gagtgt	999
aaa	taaa	.gcc	ttgt	tgat	tg a	agat	rgga	t aa	taga	accc	gra	acya	tac	taga	tatgca	1059
atg	gtct	tgg	gcaa	acat	ac c	tggt	tgta	c aa	CLLL	ayca	- ccy	gygc	taa	caac	agggta	1119
aaa	.gcta	aat	ggag	tttc	tc c	tgct	ctgt	.c ca		ctac	yaa	+++0	rage	atta	ttgaga	1179
agg	ctgg	gag	gatt	gtgt	at t	ttgc	ttca	c ay	tatt	acaa	ctt	gaca	act	ccaa	atttgc	1239
ago	gtat	ttc	actt	+	at c	ttacc	eact	a at	taat	attt	cat	cttt	tac	ttat	ttctac	1299
att	acta	taay	anta	.cayc	at a	ttta	ggaa	a to	tcac	gato	tto	aaag	gaa	aggg	taaaaa	1359
aag	ttes	taa	aaca	18880	ict o	tatt	tago	a ca	tgat	ttta	ttg	tatt	gcg	ttat	tagctg	1419
att	ttac	rtca	tttt	atat	tt c	rcaaa	ataa	ıa tt	tcta	ıatat	: tta	ittga	iaat	tgct	taattt	1479
ac =	cacc	cta	taca	caca	iga a	aato	ratat	a aa	ıatat	gaga	acc	jaagt	tta	aaat	tgtgac	1539
+ ~+	~a++	cat	tata	acac	raa d	tttta	aatt	t co	cago	tttt	: tga	agat	tta	agct	acgeta	1599
++=	artac	ttc	cctt	tato	ta t	acca	taac	it go	ttga	laaac	: gtt	:aagg	jttt	toto	getetge	1659
+++	- ~+++	-++	taat	atica	aa a	agagt	caat	g to	jaaco	cttgg	, ttc	ggaco	cca	agt	cacaag	1719
att	-+++	nnse	tgat	:dada	ac d	taca	gaca	at to	ctgcc	taga	a ttt	cacta	agcg	rgrg	JCCLLLL	1779 1839
gco	etget	tct	cttt	gatt	tc a	caga	atat	t ca	ittca	igaac	l cc	gegti	ITCT	gtag	gtgtggt	1899
gga	attco	ccac	tggg	gctct	gg t	cctt	ccct	t gg	gated	cgtc	agu	tagu	cac	ttt	agcggct	1959
tgo	cacg	taga	cttg	gctag	gga a	agaaa	tege	ig ag	Jocas	stact	- ac	arar	rcat	ttt	gcatttg	2019
gaa	actc	ttta	agco		greg d	19 cg	-220	cc ac	rtaa	atac	- ac	actt	aatt	cat	catgact	2079
70	cgcg	LCaa	gaag	tatti	tee .	ratta	agati	tt ac	ataa	ageta	a ca	catt	aata	tgt	atcgcct	2139
+=	~~~	anca	act	atati	tee .	agga	acca	ga to	cacq	attti	t ta	gcca	tgga	aca	atatatc	2199
CC	ataa	ana	aga	catti	tca	atato	gaac	ta t	tcta	ととしたり	t gt	gtta	taat	tta	aacttcg	2259
a t	++~~	tcat	agt	catt	taa	atta	acat	tt c	tgct	tacto	g ct	actg	gatt	בבב	gctgcag	2319
22	atat	atca	ata	acco	aca	ttaa	acat	ac c	agtt	ggato	c at	gata	agca	aaa	tgaaaga	2379
22	+==+	att	aad	aaaa.	aat	taaq	taac	ta t	gtta	cact	g ct	tctc	ccat	gcc	agagaat	2439 2499
aa	actc	tttc	aag	catc	atc	tttg	aaga	gt c	gtgt	ggtg	r ga	attg	gttt tast	gtg	tacatta	2559
ga	atgt	atgc	aca	catc	cat	ggac	actc	ag g	acat	agtt aatt	y gc + +~	tota	acct	cyg n+=	ggcatgg	2619
gt	aaaa	ctta	tga	aaat	TTC	ctca	tgct	ya a a+ ~	cege	aatt aatt	t 22	tata	++++	gea	aagtaaa tataatg	2679
at	ttag	atca	att	ccat	gtc	ttta	erad	yr a	atar	tatt	a us	atta	tata	gat	tctaact	2739
gg	tat9	2222	. add	tage	ctr	acct	agay	aa d	atca	ctta	a at	gtta	ggtt	tca	tagaact	2799
a t	acta	atct	+ + +	caca	aaa	aato	tata	aa a	taca	gtcg	t tg	raaaa	aaat	ttt	gtatcaa	2859
25	+~+	taas	aab	ttac	raaσ	cttc	tcct	ta a	ccta	rtatt	g at	acto	jactt	. gaa	iccacccc	2919
ct	aaaa	attaa	gag	ccgt	ata	ccta	cctg	rta a	gtct	tttc	a ca	tato	attt	aaa	cttttgt	2979

```
ttgtattatt actgatttac agcttagtta ttaatttttc tttataagaa tgccgtcgat
3099
aaatctttca ctgtctctaa tggctgtgct gtttaacatt ttttgaccct aaaattcacc
                                                                   3159
aacagtctcc cagtacataa aataggctta atgactggcc ctgcattctt cacaatattt
                                                                   3219
ttccctaagc tttgagcaaa gttttaaaaa aatacactaa aataatcaaa actgttaagc
agtatattag tttggttata taaattcatc tgcaatttat aagatgcatg gccgatgtta
                                                                   3339
atttgcttgg caattctgta atcattaagt gatctcagtg aaacatgtca aatgccttaa
                                                                   3399
attaactaag ttggtgaata aaagtgccga tctggctaac tcttacacca tacatactga
tagtttttca tatgtttcat ttccatgtga tttttaaaaat ttagagtggc aacaattttg
cttaatatgg gttacataag ctttatttt tcctttgttc ataattatat tctttgaata
                                                                   3579
ggtctgtgtc aatcaagtga tctaactaga ctgatcatag atagaaggaa ataaggccaa
                                                                   3639
gttcaagacc agcctgggca acatatcgag aacctgtcta caaaaaaatt aaaaaaaatt
                                                                   3699
agccaggcat ggtggcgtac actgagtagt ttgtcccagc tactcgggag ggtgaggtgg
                                                                   3759
gaggatcgct tcagcccagg aggttgagat tgcagtgagc catggacata ccactgcact
                                                                   3819
acagcctagg taacagcacg agaccccaac tcttagaaaa tgaaaaggaa atatagaaat
ataaaatttg cttattatag acacacagta actcccagat atgtaccaca aaaaatgtga
aaagagagag aaatgtctac caaagcagta ttttgtgtgt ataattgcaa gcgcatagta
aaataatttt aaccttaatt tgtttttagt agtgtttaga ttgaagattg agtgaaatat
                                                                   4059
tttcttggca gatattccgt atctggtgga aagctacaat gcaatgtcgt tgtagttttg
                                                                    4119
catggcttgc tttataaaca agattttttc tccctccttt tgggccagtt ttcattacga
                                                                    4179
gtaactcaca ctttttgatt aaagaacttg aaattacgtt atcacttagt ataattgaca
                                                                   4239
ttatatagag actatgtaac atgcaatcat tagaatcaaa attagtactt tggtcaaaat
                                                                   4299
atttacaaca ttcacatact tgtcaaatat tcatgtaatt aactgaattt aaaaccttca
actattatga agtgctcgtc tgtacaatcg ctaatttact cagtttagag tagctacaac
                                                                    4419
tettegatac tateateaat atttgacate ttttccaatt tgtgtatgaa aagtaaatet
                                                                    4479
attectgtag caactgggga gtcatatatg aggtcaaaga catatacett gttattataa
                                                                    4539
tatgtatact ataataatag ctggttatcc tgagcagggg aaaaggttat ttttaggaaa
                                                                    4599
accacttcaa atagaaagct gaagtacttc taatatactg agggaagtat aatatgtgga
                                                                    4659
acaaactctc aacaaaatgt ttattgatgt tgatgaaaca gatcagtttt tccatccgga
                                                                    4719
ttattattgg ttcatgattt tatatgtgaa tatgtaagat atgttctgca attttataaa
                                                                    4779
tgttcatgtc tttttttaaa aaaggtgcta ttgaaattct gtgtctccag caggcaagaa
                                                                    4839
 tacttgacta actetttttg tetetttatg gtattttcag aataaagtet gaettgtgtt
                                                                    4899
 tttgagatta ttggtgcctc attaattcag caataaagga aaatatgcat ctcaaaaat
                                                                    4958
 <210> 122
<211> 5094
 <212> DNA
 <213> Homo sapiens
 <221> misc_feature
 <222> 31..33
 <223> ATG
 <221> misc_feature
 <222> 976..978
 <223> TAG
 <221> polyA_signal
 <222> 5067..5072
 <223> AATAAA
 <400> 122
 ctgctgtccc tggtgctcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc
                                                                      54
                                 Met Arg Tyr Leu Leu Pro Ser Val
                                  1
 gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg
                                                                     102
 Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
                         15
 cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac
                                                                      150
 Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
                                         35
                     30
  gac egg etg tac tge gte tac eag age atg gtg ete tte tte tte gag
                                                                      198
 Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Phe Glu
                                     50
                  45
  aat tac acc ggg gtc cag ata ttg cta tat gga gat ttg cca aaa aat
                                                                      246
```

	-	Thr	60					65					70			
aaa Lys	gaa Glu	aat Asn 75	ata Ile	ata Ile	tat Tyr	tta Leu	gca Ala 80	aat Asn	cat His	caa Gln	agc Ser	aca Thr 85	gtt Val	gac Asp	tgg Trp	294
att Ile	Val	gct Ala	gac Asp	atc Ile	ttg Leu	gcc Ala 95	atc	agg Arg	cag Gln	aat Asn	gcg Ala 100	cta Leu	gga Gly	cat His	gtg Val	342
cgc Arg 105	90 tac Tyr	gtg Val	ctg Leu	aaa Lys	gaa Glu 110	aga	tta Leu	aaa Lys	tgg Trp	ctg Leu 115	cca	ttg Leu	tat Tyr	ggg Gly	tgt Cys 120	390
tac Tyr	Phe	gct Ala	Gln	His 125	gga Gly	Gly	Ile	Tyr	Val 130	Lys	Arg	Ser	Ala	Lys 135	Pne	438
aac Asn	gag Glu	aaa Lys	gag Glu 140	ato	cga Arg	aac Asn	aag Lys	ttg Leu 145	cag Gln	agc Ser	tac Tyr	gtg Val	gac Asp 150	gca Ala	gga Gly	486
act Thr	cca Pro	atg Met 155	tat	ctt Leu	gtg Val	att Ile	ttt Phe 160	cca Pro	gaa Glu	ggt Gly	aca Thr	agg Arg 165	tat Tyr	aat Asn	cca Pro	534
gag Glu	Gln	aca Thr	aaa Lys	gtc Val	ctt Leu	tca Ser 175	act	agt Ser	cag Gln	gca Ala	ttt Phe 180	gct Ala	gcc Ala	caa Gln	cgt Arg	582
Gly	Lys	gac Asp	gat Asp	gga Gly	ggg Gly 190	caq	cga Arg	aga Arg	gag Glu	tca Ser 195	ccg Pro	acc Thr	atg Met	acg Thr	gaa Glu 200	630
185 ttt Phe	ctc	tgc Cys	aaa Lys	gaa Glu 205	tat	cca Pro	aaa Lys	att Ile	cat His 210	att Ile	cac His	att Ile	gat Asp	cgt Arg 215	TTE	678
gac Asp	aaa Lys	aaa Lys	gat Asp 220	gtc Val	cca Pro	gaa Glu	gaa Glu	caa Gln 225	gaa Glu	cat	atg Met	aga Arg	aga Arg 230	Trp	ctg Leu	726
cat His	gaa Glu	. Arg	ttc Phe	даа	atc Ile	aaa Lys	gat Asp 240	aag Lys	atg	ctt Leu	ata Ile	gaa Glu 245	Pne	tat Tyr	gag Glu	774
tca Sei	Pro	Asp	cca	gaa Glu	aga Arg	aga Arg 255	aaa Lys	aga	ttt Phe	cct Pro	ggg Gly 260	aaa Lys	agt	gtt Val	aat Asn	822
Sei	Lys	. ++=	agt Ser	atc Ile	aag Lys 270	aag Lys	act	tta Lei	a cca ı Pro	tca Ser 275	ato Met	tto	g ato	tta Leu	agt Ser 280	870
269 gg Gl	- ++	g act ı Thi	gca Ala	a Gly	ato Met	ctt	ato Met	g acc	gat As <u>r</u> 290	gct Ala	gga	a ago 7 Aro	g aag g Lys	cto Lev 295	tat Tyr	918
gt: Va	g aad 1 Asi	c acc	r Tri	, Ile	tat	gga Gly	a aco	c cta r Lem	a cti u Lei	t gg	c tgo y Cy:	c cto	g tgg ı Trp 310	g gtt p Val	act L Thr	966
		a gca s Ala	a *	g aca	aagta	agct	gtc			agtg	ggat	gt g			t	1018
ta gt	aagc cttg taaa	cttg ggca tgga	cgg ttg aac gtt	attga atac tctc	aag a ctg a ctg a	attg gttg ctct gcaa	gata taca gtcc gtca	at a ac t at ‡ ga t	gaat ttag tcct ggct	ttgt catc atga gcat	g ac g gg a ct t tt	gaaa gctg aatg tgag	gctg ctgg acaa catt	ata aag ctt aat	gtgtaaa tgcaatg ggtaaaa gagaagg ttgcagc	1078 1138 1198 1258 1318 1378
gt ac tc	attt taaa agtg catg	cact gtat gaaat gggga	ttt tta aaa aaa	tctg gtat ttgt agct	tta ctt att ctg	tttt gcag tagg ttta aaat	caat ctag aagt gcac	tt a tt a gt c at g	ttac atat agga attt taat	aact ttca tgtt tatt	t ga t ct c aa g ta	cago tttg agga ttgc	ctta ctta aagg gtta ittgc	agc ttt gta tta tta	ctacaag aaaagtg gctgatt atttgca	1438 1498 1558 1618
C	accct	gtac	aca	caga	aaa	tggt	ataa	aa t	atga	gaac	g aa	igttt	aaaa	ttg	tgactct	1678

gattcattat	agcagaactt	taaatttccc .	agctttttga	<b>agattta</b> agc	tacgctatta	1738
atacttacat	ttatctatac	cataaqtqct	tqaaaacqtt	aaggttttct	gttttgttt	1798
attttttaa	tatcaaaaga	atcaatataa	accttggttg	gaccccaagt	tcacaagatt	1858
tttaaggtga	tgagageetg	cagacattct	gcctagattt	actagegtgt	geettttgee	1918
tacttctctt	tgatttcaca	gaatattcat	tcagaagtcg	cgtttctgta	gtgtggtgga	1978
ttcccactaa	actictaatics	ttcccttgga	tcccgtcagt	ggtgctgctc	ageggeetge	2038
acgtagactt	gctaggaaga	aatgcagagc	cagcctgtgc	tgcccacttt	cagagilgaa	2098
ctctttaagc	ccttgtgagt	gggcttcacc	agctactgca	gaggcatttt	geattigtet	2158 2218
atataaaaa	atteacette	tcaagccagt	gaaatacaga	cttaattcgt	catgactgaa	2218
casatttatt	tatttcccat	taggtttagt	ggagctacac	attaatatgi	accyccitay	2338
agcaagaget	atattccaga	aaccagatca	cgatttttag	ccatggaaca	atatateeca	2336
taaaaaaaa	cctttcagtg	tgaactgttc	tatttttgtg	ttataatiia	aacttegatt	2458
tcctcatagt	cctttaagtt	gacatttctg	cttactgcta	ctggattttt	getgeagaaa	2518
tatatcagtg	gcccacatta	aacataccag	ttggatcatg	ataagcaaaa	rgadagadac	2578
aatgattaag	ggaaaattaa	gtgactgtgt	tacactgett	Ctcccargue	tagagaacaaa	2638
ctctttcaag	catcatcttt	gaagagtcgt	gtggtgtgaa	etggtttgtg	cacatagaa	2698
tgtatgcaca	catccatgga	cactcaggat	atagttggcc	taataategg	ggcatgggta	2758
aaacttatga	aaatttcctc	atgctgaatt	gtaattttct	cttacctgta	aaytaaaatt	2818
Fagatosatt	ccatotottt	oftaaotaca	gggatttaat	atattttgaa	Lacaacygyc	2878
atattataaa	tttgaacttt	gagaggcaat	actottogaa	ttatgtggat	LULAAULUAL	2938
tttaacaacc	tagectgace	tocataagat	cacttgaatg	ttaggtttta	Lagaactata	2998
ctaatcttct	cacaaaaaggt	ctataaaata	cagtcgttga	adadadilli	graccadac	3058
gtttggaaaa	ttagaagctt	ctccttaacc	tgtattgata	ctgacttgaa	ctattttttta	3118
aaattaagag	ccgtatacct	acctgtaagt	cttttcacat	atcatttaaa	arterstata	3178
+ = + + = + + = < +	gatttacago	ttagttatta	atttttttt	ataagaatge	cgicgatgig	3238
catgctttta	tgtttttcag	aaaagggtgt	gtttggatga	aagtaaaaa	addacadaa	3298
tatttaacta	tetetaatgg	ctatactatt	taacattttt	tgaccctaaa	acceaccaac	3358
agtctcccag	tacataaaat	aggcttaatg	actggccctg	cattetteat	attaaccact	3418
cctaagcttt	gagcaaagtt	ttaaaaaaaat	acactadaal	adicadadce	getaageage	3478
atattagttt	ggttatataa	attcatctgc	aatttataag	atgeatgget	gatyttaatt	3538
tgcttggcaa	ttctgtaatc	attaagtgat	ctcagtgaaa	tacacatac	atactgatag	3598
aactaagttg	gtgaataaaa	gtgccgatct	ggetaaetet	cacaccacac	acaccyacus	3658
tttttcatat	gtttcatttc	catgtgattt	ttaaaattta	gagtggtaat	ttaaataaat	3718
aatatgggtt	acataagctt	tatttttcc	tttgttcata	actatatte	aggccaagtt	3778
ctgtgtcaat	caagtgatct	aactagactg	atcatagata	gaaggaaaca	aaaaattacc	3838
caagaccago	ctgggcaaca	tatcgagaac	ctgtctacaa	tagggaggg	gaggaggg	3898
caggcatggt	ggcgtacact	gagtagtttg	teceagetae	ccgggagggc	ctocactaca	3958
gatcgcttca	a gcccaggagg	ttgagattgc	agtgagccat	ggacacacca	tagaaatata	4018
gcctaggtaa	a cagcacgaga	ccccaactct	tagaaaatga	taggaaaca	aatotoaaaa	4078
aaatttgctt	attatagaca	cacagtaact	cccagataty	attacaaaa	catactaaaa	4138
gagagagaa	a tgtctaccaa	agcagtattt	tgtgtgtata	accycaaycy	caeatattt	4198
taattttaad	ctraatttgt	ttttagtagt	gtttagattg	adjactgag	agttttgcat	4258
cttggcagai	attccgtatc	tggtggaaag	ctacaatyca	acgeograps	attacgagta	4318
ggcttgctt	t ataaacaaga		ttaccttatc	acttantata	attgacatta	4378
actcacact	t tttgattaaa t atgtaacatg	gaacttgaaa	cacyclact	actedgedes	tcaaaatatt	4438
tatagagac	t atgtaacatg	caatcattay	tataattaac	tgaatttaaa	accttcaact	4498
tacaacatt	c acatacttgt t gctcgtctgt	cadatattca	. Lytaactaac	tttagagtag	ctacaactct	4558
attatgaag	t getegtetgt t catcaatatt	. acaategeta	tocaatttot	r ctatcaaaa	taaatctatt	4618
tcgatacta	t catcaatati a ctggggagto	tgacatette	tcaacccgt	atacettett	attataatat	4678
cctgtagca	a ctggggagt	atatatgay	ccaaagaca	accettattt	taggaaaacc	4738
gtatactat	a ataatagete	gulateerge	tatactgag	r gaagtataat	atgtggaaca	4798
acttcaaat	a gaaagetgaa	trastatta	t deacegage	cagtttttc	atccggatta	4858
aactctcaa	c aaaatgttta	tigatgitgi	r cgaaacaga.	ttctgcaati	ttataaatgt	4918
ttattggtt	c acgattta	. acycyaacad	r aaattototo	tctccagca	g gcaagaatac	4978
tcatgtctt	t atteteate	- otttstatt	tittcagaa	t aaagtctga	c ttgtgtttt	5038
ttgactaac	g gtgcctcat	t sattcaces	taaaggaaa	a tatocatct	c aaaaat	5094
		. autocayout		<b></b>		
<210> 123 <211> 504						
<211> 504 <212> DNA						
	no sapiens					
~~±J~ HOI	CARDTERNO CO					

```
<220>
<221> misc_feature
<222> 31..33
<223> ATG
<221> misc_feature
<222> 931..933
<223> TAG
<221> polyA_signal
<222> 5022..5D27
<223> AATAAA
<400> 123
ctgctgtccc tggtgctcca cacgtactcc atg cgc tac ctg ctg ccc agc gtc
                                 Met Arg Tyr Leu Leu Pro Ser Val
                                 1
gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg
                                                                     102
Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
                        15
egg etg etc tec gec tte etg ecc gec ege tte tac caa geg etg gac
                                                                     150
Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
                    30
gac egg etg tac tge gte tac eag age atg gtg etc tte tte tte gag
                                                                     198
Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Glu
                                    50
                45
aat tac acc ggg gtc cag ata ttg cta tat gga gat ttg cca aaa aat
                                                                      246
Asn Tyr Thr Gly Val Gln Ile Leu Leu Tyr Gly Asp Leu Pro Lys Asn
                                65
            60
aaa gaa aat ata ata tat tta gca aat cat caa agc aca gtt gac tgg
                                                                      294
Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr Val Asp Trp
                                                85
                            80
att gtt gct gac atc ttg gcc atc agg cag aat gcg cta gga cat gtg
                                                                      342
Ile Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu Gly His Val
                                            100
                         95
ege tac gtg ctg aaa gaa ggg tta aaa tgg ctg cca ttg tat ggg tgt
                                                                      390
Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys
                                        115
                    110
 tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt
                                                                      438
 Tyr Phe Ala Gin His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe
                                                        135
                                     130
                125
 aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga
                                                                      486
 Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly
                                 145
             140
 act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aat cca
                                                                      534
 Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro
                                                165
                             160
 gag caa aca aaa gtc ctt tca gct agt cag gca ttt gct gcc caa cgt
                                                                      582
 Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg
                         175
                                            180
 gaa ttt ctc tgc aaa gaa tgt cca aaa att cat att cac att gat cgt
                                                                      630
 Glu Phe Leu Cys Lys Glu Cys Pro Lys Ile His Ile His Ile Asp Arg
                                         195
                    190
 atc gac aaa aaa gat gtc cca gaa gaa caa gaa cat atg aga aga tgg
                                                                       678
 Ile Asp Lys Lys Asp Val Pro Glu Glu Glu His Met Arg Arg Trp
                                     210
                 205
 ctg cat gaa cgt ttc gaa atc aaa gat aag atg ctt ata gaa ttt tat
                                                                       726
 Leu His Glu Arg Phe Glu Ile Lys Asp Lys Met Leu Ile Glu Phe Tyr
                                 225
             220
  gag tca cca gat cca gaa aga aga aaa aga ttt cct ggg aaa agt gtt
                                                                       774
 Glu Ser Pro Asp Pro Glu Arg Arg Lys Arg Phe Pro Gly Lys Ser Val
                            240
                                                 245
  aat tee aaa tta agt ate aag aag aet tta eea tea atg ttg ate tta
                                                                       822
  Asn Ser Lys Leu Ser Ile Lys Lys Thr Leu Pro Ser Met Leu Ile Leu
```

78

200	
250 255 260 agt ggt ttg act gca ggc atg ctt atg acc gat gct gga agg aag ctg	870
Ser Gly Leu Thr Ala Gly Met Leu Met Thr Asp Ala Gly Arg Lys Leu	
265 270 275 280	
tat gra age acc tog ata tat gga acc cta ctt ggc tgc ctg tgg gtt	918
Tyr Val Asn Thr Trp Ile Tyr Gly Thr Leu Leu Gly Cys Leu Trp Val	
285 290 293	073
act att aaa gca tag acaagtagct gtctccagac agtgggatgt gctacattgt	973
Thr Ile Lys Ala *	
300	1033
ctatttttgg cggctgcaca tgacatcaaa ttgtttcctg aatttattaa ggagtgtaaa taaagccttg ttgattgaag attggataat agaatttgtg acgaaagctg atatgcaatg	1093
gtettgggca aacatacetg gttgtacaac tttagcateg gggetgetgg aagggtaaaa	1153
gctaaatgga gtttctcctg ctctgtccat ttcctatgaa ctaatgacaa cttgagaaagg	1213
ctgggaggat tgtgtatttt gcaagtcaga tggctgcatt tttgagcatt aatttgcagc	1273
grattroact tritcigtta tittcaatti attacaacti gacagcicca agcicitati	1333
actagagrat tragtatett geagetagtt aatattteat ettitgetta titetaeaag	1393
tragtogaat agattotatt taggaagtot caggatottc aaaggaaagg gidaddayig	1453
treateggga aggactetg titageacat gattitattg tattgegtta tragergati	1513
ttactcattt tatatttgca aaataaattt ctaatattta ttgaaattgc ttaatttgca	1573
caccetgrac acacagaaaa tggtataaaa tatgagaacg aagtttaaaa tegtgactet	1633
gattcattat agcagaactt taaatttccc agctttttga agatttaagc tacgctatta	1693 1753
gtacttccct tratctatac cataagtact tgaaaacgtt aaggttttct gttttgtttt	1813
gettettea tatcaaaaga gecggtgtga accettggteg gaccccaage ccacaagace	1873
tttaaggtga tgagagcctg cagacattct gcctagattt actagcgtgt gccttttgcc	1933
tgcttctctt tgatttcaca gaatattcat tcagaagtcg cgtttctgta gtgtggtgga	1993
ttcccactgg gctctggtcc ttcccttgga tcccgtcagt ggtgctgctc agcggcttgc acgtagactt gctaggaaga aatgcagagc cagcctgtgc tgcccacttt cagagttgaa	2053
acgtagactt gctaggaaga datgcagagt tagtetgtgt tagtetattt gcatttgtct ctctttaagc ccttgtgagt gggcttcacc agctactgca gaggcatttt gcatttgtct	2113
gtgtcaagaa gttcaccttc tcaagccagt gaaatacaga cttaattcgt catgactgaa	2173
cgaatttgtt tatttcccat taggtttagt ggagctacac attaatatgt atcgccttag	2233
accapaget stattcagg aaccagatca cgatttttag ccatggaaca atatatcca	2293
taggagaga cottoggto togactotto tatttttgtg ttataattid dacticyatt	2353
toctcatagt cotttaagtt gacatttotg ottactgota otggattict gotgoagaa	2413
tatatcadto occapatta aacataccad tiddatcatg atadgcadad tydaayadat	2473
astrattaan graaaattaa gtgactgtgt tacactgctt ctcccatgcc agagaataaa	2533 2593
crettreage cateatettt gaagagtegt gtggtgtgaa ttggtttgtg tacattagaa	2653
tgtatgcaca catccatgga cactcaggat atagttggcc taataatcgg ggcatgggta	2713
aaacttatga aaatttcctc atgctgaatt gtaattttct cttacctgta aagtaaaatt	2773
tagatcaatt ccatgtcttt gttaagtaca gggatttaat atattttgaa tataatgggt	2833
atgitetaaa titgaactit gagaggcaat actgitggaa tiatgiggat tetaacteat	2893
tttaacaagg tagcctgacc tgcataagat cacttgaatg ttaggtttca tagaactata ctaatcttct cacaaaaggt ctataaaata cagtcgttga aaaaaatttt gtatcaaaat	2953
gtttggaaaa ttagaagett eteettaace tgtattgata etgaettgaa ttattteeta	3013
aaattaagag ccgtatacct acctgtaagt cttttcacat atcatttaaa cttttgtttg	3073
tattattact datttacade ttadttatta attitititit ataagaatge egitegatgeg	3133
catgotttta tottittoag aaaagggtgt gtttggatga aagtaaaaaa aaaaacadaa	3193
totttoacto tototaatoo otgtoctott taacattttt tgaccctaaa atteaceade	3253
agtotocoag tacataggat aggottaatg actggccctg cattcttcac datattitt	3313
cotaagettt gaggaaagtt ttaaaaaaaat acactaaaat aatcaaaact gttaageagt	3373 3433
atattagttt ggttatataa attcatctgc aatttataag atgcatggcc galgulaalu	3433
trottograa trotograato attaagtoat otcagtoaaa catotcaaat goottaaatt	3553
aactaagttg gtgaataaaa gtgccgatct ggctaactct tacaccatac atactgatag	3613
ttttcatat gtttcatttc catgtgattt ttaaaattta gagtggcaac aattttgctt	3673
aatatgggtt acataagctt tattttttcc tttgttcata attatattct ttgaataggt	3733
ctgtgtcaat caagtgatct aactagactg atcatagata gaaggaaata aggccaagtt caagaccagc ctgggcaaca tatcgagaac ctgtctacaa aaaaattaaa aaaaattagc	3793
caagaccage etgggcaaca tategagaae etgeetacaa addatedaa aaggetgggag caggcatggt ggcgtacact gagtagtttg teccagetae tegggagggt gaggtgggag	3853
gatogottoa googaggagg ttgagattgc agtgagccat ggacatacca ctgcactaca	3913
gostaggtaa cagcacgaga ccccaactct tagaaaatga aaaggaaata tagaaatata	3973
aaatttgctt attatagaca cacagtaact cccagatatg taccacaaaa aatgtgaaaa	4033

```
gagagagaaa tgtctaccaa agcagtattt tgtgtgtata attgcaagcg catagtaaaa
                                                                     4093
taattttaac cttaatttgt ttttagtagt gtttagattg aagattgagt gaaatatttt
                                                                     4153
cttggcagat attccgtatc tggtggaaag ctacaatgca atgtcgttgt agttttgcat
                                                                     4213
ggettgettt ataaacaaga ttttttetee eteettttgg geeagtttte attacgagta
                                                                     4273
actcacactt tttgattaaa gaacttgaaa ttacgttatc acttagtata attgacatta
                                                                     4333
tatagagact atgraacatg caatcattag aatcaaaatt agtactttgg tcaaaatatt
                                                                     4393
tacaacattc acatacttgt caaatattca tgtaattaac tgaatttaaa accttcaact
                                                                     4453
attatgaagt gctcgtctgt acaatcgcta atttactcag tttagagtag ctacaactct
                                                                     4513
togatactat catcaatatt tgacatottt tocaatttgt gtatgaaaag taaatotatt
                                                                     4573
cctgtagcaa ctggggagtc atatatgagg tcaaagacat ataccttgtt attataatat
gtatactata ataatagctg gttatcctga gcaggggaaa aggttatttt taggaaaacc
                                                                     4693
acttcaaata gaaagctgaa gtacttctaa tatactgagg gaagtataat atgtggaaca
                                                                     4753
aactctcaac aaaatgttta ttgatgttga tgaaacagat cagtttttcc atccggatta
                                                                     4813
ttattggttc atgattttat atgtgaatat gtaagatatg ttctgcaatt ttataaatgt
                                                                     4873
tcatgtcttt ttttaaaaaa ggtgctattg aaattctgtg tctccagcag gcaagaatac
                                                                     4933
ttgactaact ctttttgtct ctttatggta ttttcagaat aaagtctgac ttgtgttttt
                                                                     4993
gagattattg gtgcctcatt aattcagcaa taaaggaaaa tatgcatctc aaaaat
                                                                     5049
<210> 124
<211> 5324
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<222> 31..33
<223> ATG
 <221> misc_feature
 <222> 586..588
 <223> TAA
 <221> polyA_signal 
<222> 5297..5302
 <223> AATAAA
 <400> 124
 etgetgtece tggtgeteca caegtactee atg ege tae etg etg ece age gte
                                                                        54
                                  Met Arg Tyr Leu Leu Pro Ser Val
                                   1
 gtg ctc ctg ggc acg gcg ccc acc tac gtg ttg gcc tgg ggg gtc tgg
                                                                       102
 Val Leu Leu Gly Thr Ala Pro Thr Tyr Val Leu Ala Trp Gly Val Trp
                         15
 cgg ctg ctc tcc gcc ttc ctg ccc gcc cgc ttc tac caa gcg ctg gac
                                                                       150
 Arg Leu Leu Ser Ala Phe Leu Pro Ala Arg Phe Tyr Gln Ala Leu Asp
                                          35
                      30
 gac egg etg tae tge gte tae eag age atg gtg ete tte tte tgag
                                                                       198
 Asp Arg Leu Tyr Cys Val Tyr Gln Ser Met Val Leu Phe Phe Glu
                                      50
                  45
 aat tac acc ggg gtc cag ata ttg cta tat gga gat ttg cca aaa aat
                                                                       246
 Asn Tyr Thr Gly Val Gln Ile Leu Leu Tyr Gly Asp Leu Pro Lys Asn
                                  65
              60
  aaa gaa aat ata ata tat tta gca aat cat caa agc aca gtt gac tgg
                                                                        294
 Lys Glu Asn Ile Ile Tyr Leu Ala Asn His Gln Ser Thr Val Asp Trp
                             80
  att gtt gct gac atc ttg gcc atc agg cag aat gcg cta gga cat gtg
                                                                        342
  Ile Val Ala Asp Ile Leu Ala Ile Arg Gln Asn Ala Leu Gly His Val
                                              100
                          95
  cgc tac gtg ctg aaa gaa ggg tta aaa tgg ctg cca ttg tat ggg tgt
                                                                        390
  Arg Tyr Val Leu Lys Glu Gly Leu Lys Trp Leu Pro Leu Tyr Gly Cys
                                                               120
                                           115
                      110
  105
  tac ttt gct cag cat gga gga atc tat gta aag cgc agt gcc aaa ttt
                                                                        438
  Tyr Phe Ala Gln His Gly Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe
                                      130
                  125
  aac gag aaa gag atg cga aac aag ttg cag agc tac gtg gac gca gga
                                                                        486
  Asn Glu Lys Glu Met Arg Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly
```

150	
140 145 150	534
act cca atg tat ctt gtg att ttt cca gaa ggt aca agg tat aat cca	224
Thr Pro Met Tyr Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro	
gag caa aca aaa gtc ctt tca gct agt cag gca ttt gcc caa cgt	582
Glu Gln Thr Lys Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg	
170 175 160 ggc taa agcagtcctc ctgagtagtt aggactacag acatacacgt gccaccgcgc	638
cageteegt gttetetttg ttteeetgee teetgetett ceaettatet ttgeatggea	698
recettered tattaaaaca tetertaaca ccacgaataa aggedaetta cyttyettet	758
cattagataa agaattatti agatgcaatt tatgatgtta cggtggttta tydayyydda	818
gargaturan nucancuaan adagteaced accatgacgg aatttetety cadayaacyt	878
gananaatto atattoacat toatcotato qacaaaaaag atgtcccaga agaacaagaa	938
antatgagaa gatggctgca tgaacgtttc gaaatcaaag ataagatgct tatagaattt	998
tatanatana cadatcada aadaadaaaa adatttcctg ggadaayigi taatteedaa	1058
thangtates agaagactt accatcaatg tigatettaa giggitigae igeaggeatg	1118 1178
attatgacca atactagaag gaagctgtat gtgaacacct ggatatalgg daccetactt	1238
gastacetat adattactat taaagcatag acaagtaget gteteeagae agtgggaege	1298
actionstat ctattttag cogctgcaca tgacatcaaa ttgtttcccg dattattaa	1358
ggagtgtaaa taaagccttg ttgattgaag attggataat agaatttgtg acgaaagctg	1418
atatgcaatg gtcttgggca aacatacctg gttgtacaac tttagcatcg gggctgctgg	1478
aagggtaaaa gctaaatgga gtttctcctg ctctgtccat ttcctatgaa ctaatgacaa aagggtaaaa gctaaatgga gtttctcctg ctctgtccat ttcctatgaa ctaatgacaa	1538
cttgagaagg ctgggaggat tgtgtatttt gcaagtcaga tggctgcatt tttgagcatt aatttgcagc gtatttcact ttttctgtta ttttcaattt attacaactt gacagctca	1598
agetettatt actaaagtat ttagtatett geagetagtt aatatteet ettetgetta	1658
tttctacaag tcagtgaaat aaattgtatt taggaagtgt caggatgttc aaaggaaagg	1718
gtaaaaagtg ttcatgggga aaaagctctg tttagcacat gattttattg tattgcgtta	1778
ttagctgatt ttactcattt tatatttgca aaataaattt ctaatattta ttgaaattgc	1838
ttaatttgca caccetgtac acacagaaaa tggtataaaa tatgagaacg aagtttaaaa	1898
ttataactot dattoattat addadaactt taaatttccc agctttttga agatttaage	1958
taggetatta gracticeet tigicigige cataagiget igaadaegit aaggittiet	2018
getttgettt gettetttaa tatcaaaaga geoggegega acceeggeeg gacceeaage	2078
toacaagett tttaaggtga tgagagggtg cagagattct gcctagattt actagegege	2138
mantitudes testificate teatificate gaataticat teagaagueg egittetigea	2198
gtgtggtgga ttcccactgg gctctggtcc ttcccttgga tcccgtcagt ggtgctgctc	2258
aggregation acceptangent inchangeage aathoraged Cageougus Cycocacte	2318 2378
anguattan ctotttange cottataggt aggetteace agetaciged yayyearer	2378
	2438
	2558
atcgccttag agcaagagct gtgttccagg aaccagatca cgatttttag ccatggaaca	2618
attaccca tgggagaaga cettteagtg tgaactgtte tatttttgtg ttataattta	2678
acticgatt tecteatagt cetttaagtt gaeatttetg ettactgeta etggattttt	2738
gctgcagaaa tatatcagtg gcccacatta aacataccag ttggatcatg ataagcaaaa tgaaagaaat aatgattaag ggaaaattaa gtgactgtgt tacactgct ctcccatgcc	2798
agagaataaa ctcttcaag catcatcttt gaagagtcgt gtggtgtgaa ttggtttgtg	2858
agagaataaa ctctttcaag catcatctt gaagagtege geggegege tasataatcgg tacattagaa tgtatgcaca catccatgga cactcaggat atagttggcc taataatcgg	2918
ggcatggta aaacttatga aaatttcctc atgctgaatt gtaattttct cttacctgta	2978
and the same transfer and the contractity of the transfer of the contraction of the contr	3038
tataataaan atattotaaa titgaactti gagaggcaat actgutggad udduguggad	3098
totangerest tetangenerg taggetgace togataagat cactegadeg coaggeteen	3158
tagaagtata otaatettet cacaaaaggt etataaaata cagtegiiga addaaciii	3218
That annual attendada transancti ciccitaacc intatigata cigacityaa	3278
the the anathagas contatacet acceptagge celected accaectada	3338
	3398
	3458
annotanno totttoacto tototaatoo ctotoctott taacattiil tyaccetaaa	3518 3578
	3638
This contract the decrease the translation of the second decrease	3698
gttaagcagt atattagttt ggttatataa attcatctgc aatttataag atgcatggcc	3030

3758

3818

3878 3938

3998

4058

4118

4178

4238

4298

4358

4538

4598

4658

4718

4778

4838

4958

5018

5078

5138

5198

5258 5318

5324

```
gatgttaatt tgcttggcaa ttctgtaatc attaagtgat ctcagtgaaa catgtcaaat
gccttaaatt aactaagttg gtgaataaaa gtgccgatct ggctaactct tacaccatac
atactgatag tttttcatat gtttcatttc catgtgattt ttaaaattta gagtggcaac
aattttgctt aatatgggtt acataagctt tattttttcc tttgttcata attatattct
ttgaataggt ctgtgtcaat caagtgatct aactagactg atcatagata gaaggaaata
aggecaagtt caagaccage etgggcaaca tategagaac etgtetacaa aaaaattaaa
aaaaattagc caggcatggt ggcgtacact gagtagtttg tcccagctac tcgggagggt
gaggtgggag gatcgcttca gcccaggagg ttgagattgc agtgagccat ggacatacca
ctgcactaca gcctaggtaa cagcacgaga ccccaactct tagaaaatga aaaggaaata
tagaaatata aaatttgctt attatagaca cacagtaact cccagatatg taccacaaaa
aatgtgaaaa gagagagaaa tgtctaccaa agcagtattt tgtgtgtata attgcaagcg
catagtaaaa taattttaac cttaatttgt ttttagtagt gtttagattg aagattgagt
gaaatatttt cttggcagat attccgtatc tggtggaaag ctacaatgca atgtcgttgt
agttttgcat ggcttgcttt ataaacaaga ttttttctcc ctccttttgg gccagttttc
attacgagta actcacactt tttgattaaa gaacttgaaa ttacgttate acttagtata
attgacatta tatagagact atgtaacatg caatcattag aatcaaaatt agtactttgg
tcaaaatatt tacaacattc acatacttgt caaatattca tgtaattaac tgaatttaaa
accttcaact attatgaagt gctcgtctgt acaatcgcta atttactcag tttagagtag
ctacaactct togatactat catcaatatt tgacatcttt tocaatttgt gtatgaaaag
taaatctatt cctgtagcaa ctggggagtc atatatgagg tcaaagacat ataccttgtt
attataatat gtatactata ataatagctg gttatcctga gcaggggaaa aggttatttt
taggaaaacc acttcaaata gaaagctgaa gtacttctaa tatactgagg gaagtataat
atgtggaaca aactctcaac aaaatgttta ttgatgttga tgaaacagat cagtttttcc
atccggatta ttattggttc atgattttat atgtgaatat gtaagatatg ttctgcaatt
ttataaatgt tcatgtcttt ttttaaaaaa ggtgctattg aaattctgtg tctccagcag
gcaagaatac ttgactaact ctttttgtct ctttatggta ttttcagaat aaagtctgac
ttgtgttttt gagattattg gtgcctcatt aattcagcaa taaaggaaaa tatgcatctc
aaaaat
<210> 125
<211> 77
<212> PRT
<213> Homo sapiens
<400> 125
Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
                                    10
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
                               25
            20
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
                                                45
                           40
        35
 Ser Met Val Leu Phe Phe Glu Asn Tyr Thr Gly Val Gin Leu Thr
                      55
 Gly Leu Leu Leu Thr Ser Trp Pro Ser Gly Arg Met Arg
                     70
 65
 <210> 126
 <211> 238
 <212> PRT
 <213> Homo sapiens
 <221> SITE
 <222> 98..103
 <223> Box II
 <400> 126
 Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
                        , 10
 Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
                                 25
 Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
         35
                             40
 Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln His Gly
                                             60
                       55
 Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg
```

PCT/IB98/02133 WO 99/32644 82

```
75
                 70
65
Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly Thr Pro Met Tyr Leu Val
                              90
         . 85
Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro Glu Gln Thr Lys Val Leu
                          105
         100
Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg Glu Phe Leu Cys Lys Glu
                                125
      115 120
Cys Pro Lys Ile His Ile His Ile Asp Arg Ile Asp Lys Lys Asp Val
           135 140
Pro Glu Glu Glu His Met Arg Arg Trp Leu His Glu Arg Phe Glu
                        155
      150
Ile Lys Asp Lys Met Leu Ile Glu Phe Tyr Glu Ser Pro Asp Pro Glu
                     170 175
           165
Arg Arg Lys Arg Phe Pro Gly Lys Ser Val Asn Ser Lys Leu Ser Ile
          180
                        185
Lys Lys Thr Leu Pro Ser Met Leu Ile Leu Ser Gly Leu Thr Ala Gly
                       200
                                        205
Met Leu Met Thr Asp Ala Gly Arg Lys Leu Tyr Val Asn Thr Trp Ile
                            220
  210 215
Tyr Gly Thr Leu Leu Gly Cys Leu Trp Val Thr Ile Lys Ala
225
                230
<210> 127
<211> 291
<212> PRT
<213> Homo sapiens
<221> SITE
<222> 98..103
<223> Box II
<221> SITE
<222> 149..157
<223> Box III
<400> 127
Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
                            10
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
                          25
        20
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
                  40
 Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln His Gly
                                   60
        55
 Gly Ile Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg
                                75
                 70
 Asn Lys Leu Gln Ser Tyr Val Asp Ala Gly Thr Pro Met Tyr Leu Val
                               90
              85
 Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro Glu Gln Thr Lys Val Leu
                                             110
                           105
 Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg Gly Leu Ala Val Leu Lys
                                         125
                        120
       115
 His Val Leu Thr Pro Arg Ile Lys Ala Thr His Val Ala Phe Asp Cys
                                      140
                     135
 Met Lys Asn Tyr Leu Asp Ala Ile Tyr Asp Val Thr Val Val Tyr Glu
                          155
      150
 Gly Lys Asp Asp Gly Gly Gln Arg Arg Glu Ser Pro Thr Met Thr Glu
            165 , 170
                                                 175
 Phe Leu Cys Lys Glu Cys Pro Lys Ile His Ile His Ile Asp Arg Ile
                            185
     180
 Asp Lys Lys Asp Val Pro Glu Glu Gln Glu His Met Arg Arg Trp Leu
       195 200
                                         205
 His Glu Arg Phe Glu Ile Lys Asp Lys Met Leu Ile Glu Phe Tyr Glu
                                      220
                      215
```

```
Ser Pro Asp Pro Glu Arg Arg Lys Arg Phe Pro Gly Lys Ser Val Asn
                                235
                230
Ser Lys Leu Ser Ile Lys Lys Thr Leu Pro Ser Met Leu Ile Leu Ser
                     250
             245
Gly Leu Thr Ala Gly Met Leu Met Thr Asp Ala Gly Arg Lys Leu Tyr
                                270
                265
       260
Val Asn Thr Trp Ile Tyr Gly Thr Leu Leu Gly Cys Leu Trp Val Thr
                      280
                                       285
   275
Ile Lys Ala
 290
<210> 128
<211> 261
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 68..73
<223> Box II
<221> SITE
<222> 119..127
<223> Box III
<400> 128
Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
1 5
                        10
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
                          25
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
                                      45
                      40
  35
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Met Tyr
                                    60
                  55
Leu Val Ile Phe Pro Glu Gly Thr Arg Tyr Asn Pro Glu Gln Thr Lys
                               75
                 70
Val Leu Ser Ala Ser Gln Ala Phe Ala Ala Gln Arg Gly Leu Ala Val
                             90 95
             85
Leu Lys His Val Leu Thr Pro Arg Ile Lys Ala Thr His Val Ala Phe
        100 105 110
 Asp Cys Met Lys Asn Tyr Leu Asp Ala Ile Tyr Asp Val Thr Val Val
                                       125
  115
                     120
 Tyr Glu Gly Lys Asp Asp Gly Gly Gln Arg Arg Glu Ser Pro Thr Met
                                  140
           135
 Thr Glu Phe Leu Cys Lys Glu Cys Pro Lys Ile His Ile His Ile Asp
 145 150
                                 155
 Arg Ile Asp Lys Lys Asp Val Pro Glu Glu Gln Glu His Met Arg Arg
             165 170
                                             175
 Trp Leu His Glu Arg Phe Glu Ile Lys Asp Lys Met Leu Ile Glu Phe
                                   190
                           185
          180
 Tyr Glu Ser Pro Asp Pro Glu Arg Arg Lys Arg Phe Pro Gly Lys Ser
      195 7 200 205
 Val Asn Ser Lys Leu Ser Ile Lys Lys Thr Leu Pro Ser Met Leu Ile
                            220
   210 215
 Leu Ser Gly Leu Thr Ala Gly Met Leu Met Thr Asp Ala Gly Arg Lys
 225 230 235
 Leu Tyr Val Asn Thr Trp Ile Tyr Gly Thr Leu Leu Gly Cys Leu Trp
           245 , 250
 Val Thr Ile Lys Ala
           260
 <210> 129
 <211> 90
 <212> PRT
 <213> Homo sapiens
 <400> 129
```

```
Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
                                25
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
                           40
                                               45
       35
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Asn Phe
                                          60
                       55
Ser Ala Lys Asn Val Gln Lys Phe Ile Phe Thr Leu Ile Val Ser Thr
                   70
Lys Lys Met Ser Gln Lys Asn Lys Asn Ile
                85
<210> 130
<211> 68
<212> PRT
<213> Homo sapiens
<400> 130
Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
            20
                                25
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
                           40
                                               45
        35
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Asp Ala
                       55
Tyr Arg Ile Leu
65
<210> 131
<211> 66
<212> PRT
<213> Homo sapiens
<400> 131
Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
                                25
            20
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
                                               45
                           40
  35
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Arg Leu
                       55
  50
Asp Ser
65
<210> 132
 <211> 97
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> SITE
 <222> 81..83
 <223> Box I
 <400> 132
 Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
                                    10
 Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
                                 25,
             20
 Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
                             40
         35
 Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
                                             60
                        55
 Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala
                                         75
                     70
```

```
Asn His Gln Ser Thr Asp Val Ser Cys Asp Phe Ser Arg Arg Tyr Lys
                                   90
Val
<210> 133
<211> 182
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 81..83
<223> Box I
<400> 133
Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
                                  10
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
                               25
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
                           40
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
                        55
Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala
                                      75
                   70
Asn His Gln Ser Thr Val Asp Trp Ile Val Ala Asp Ile Leu Ala Ile
                                  90
               85
Arg Gln Asn Ala Leu Gly His Val Arg Tyr Val Leu Lys Glu Gly Leu
                                                 110
                               105
         100
Lys Trp Leu Pro Leu Tyr Gly Cys Tyr Phe Ala Gln His Gly Gly Ile
                                              125
                           120
        115
Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg Asn Lys
                                       140
                       135
Leu Gln Ser Tyr Val Asp Ala Gly Thr Pro Asn Phe Ser Ala Lys Asn
                                      155
                150
Val Gln Lys Phe Ile Phe Thr Leu Ile Val Ser Thr Lys Lys Met Ser
            165
                                  170
Gln Lys Asn Lys Asn Ile
            180
<210> 134
<211> 315
 <212> PRT
 <213> Homo sapiens
<220>
 <221> SITE
 <222> 81..83
 <223> Box I
 <221> SITE
 <222> 160..165
 <223> Box II
 <400> 134
 Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
                                   10
 Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
                                25
           20
 Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
                           40
 Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
                         55
 Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala
                 70
 Asn His Gln Ser Thr Val Asp Trp Ile Val Ala Asp Ile Leu Ala Ile
                                  90
 Arg Gln Asn Ala Leu Gly His Val Arg Tyr Val Leu Lys Glu Gly Leu
```

```
105
          100
Lys Trp Leu Pro Leu Tyr Gly Cys Tyr Phe Ala Gln His Gly Gly Ile
           120
                                125
   115
Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg Asn Lys
                                      140
              135
Leu Gln Ser Tyr Val Asp Ala Gly Thr Pro Met Tyr Leu Val Ile Phe
       150
                                   155
Pro Glu Gly Thr Arg Tyr Asn Pro Glu Gln Thr Lys Val Leu Ser Ala
                               170
             165
Ser Gln Ala Phe Ala Ala Gln Arg Gly Lys Asp Asp Gly Gln Arg
                                             190
                            185
          180
Arg Glu Ser Pro Thr Met Thr Glu Phe Leu Cys Lys Glu Cys Pro Lys
                                  205
                        200
Ile His Ile His Ile Asp Arg Ile Asp Lys Lys Asp Val Pro Glu Glu
                                       220
                    215
Gln Glu His Met Arg Arg Trp Leu His Glu Arg Phe Glu Ile Lys Asp
                                   235
                  230
Lys Met Leu Ile Glu Phe Tyr Glu Ser Pro Asp Pro Glu Arg Arg Lys
                              250
              245
Arg Phe Pro Gly Lys Ser Val Asn Ser Lys Leu Ser Ile Lys Lys Thr
                         265 270
        260
Leu Pro Ser Met Leu Ile Leu Ser Gly Leu Thr Ala Gly Met Leu Met
   275 280
                                 285
Thr Asp Ala Gly Arg Lys Leu Tyr Val Asn Thr Trp Ile Tyr Gly Thr
                  295 300
Leu Leu Gly Cys Leu Trp Val Thr Ile Lys Ala
               310
305
<210> 135
<211> 300
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 81..83
<223> Box I
<221> SITE
<222> 160..165
<223> Box II
<400> 135
Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
                              10
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
                            25
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
                         40
 Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
                      55
                                      60
 Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala
                70
 Asn His Gln Ser Thr Val Asp Trp Ile Val Ala Asp Ile Leu Ala Ile
                                90
              85
 Arg Gln Asn Ala Leu Gly His Val Arg Tyr Val Leu Lys Glu Gly Leu
                                               110
                            105
          100
 Lys Trp Leu Pro Leu Tyr Gly Cys Tyr Phe Ala Gln His Gly Gly Ile
                                           125
                          120
        115
 Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg Asn Lys
                      135
                                        140
 Leu Gln Ser Tyr Val Asp Ala Gly Thr Pro Met Tyr Leu Val Ile Phe
                                    155
                150
 Pro Glu Gly Thr Arg Tyr Asn Pro Glu Gln Thr Lys Val Leu Ser Ala
                                170
```

```
Ser Gln Ala Phe Ala Ala Gln Arg Glu Phe Leu Cys Lys Glu Cys Pro
                             185
           180
Lys Ile His Ile His Ile Asp Arg Ile Asp Lys Lys Asp Val Pro Glu
                                     205
                         200
     195
Glu Gln Glu His Met Arg Arg Trp Leu His Glu Arg Phe Glu Ile Lys
                     215
                                        220
Asp Lys Met Leu Ile Glu Phe Tyr Glu Ser Pro Asp Pro Glu Arg Arg
                  230
                                     235
Lys Arg Phe Pro Gly Lys Ser Val Asn Ser Lys Leu Ser Ile Lys Lys
                                                   255
              245
                                 250
Thr Leu Pro Ser Met Leu Ile Leu Ser Gly Leu Thr Ala Gly Met Leu
                                              270
                             265
         260
Met Thr Asp Ala Gly Arg Lys Leu Tyr Val Asn Thr Trp Ile Tyr Gly
                       280
                                   285
      275
Thr Leu Leu Gly Cys Leu Trp Val Thr Ile Lys Ala
           295
<210> 136
<211> 185
<212> PRT
<213> Homo sapiens
<220>
<221> SITE
<222> 81..83
<223> Box I
<221> SITE
<222> 160..165
<223> Box II
<400> 136
Met Arg Tyr Leu Leu Pro Ser Val Val Leu Leu Gly Thr Ala Pro Thr
                                  10
Tyr Val Leu Ala Trp Gly Val Trp Arg Leu Leu Ser Ala Phe Leu Pro
                                                30
                              25
Ala Arg Phe Tyr Gln Ala Leu Asp Asp Arg Leu Tyr Cys Val Tyr Gln
                                             45
                          40
       35
Ser Met Val Leu Phe Phe Phe Glu Asn Tyr Thr Gly Val Gln Ile Leu
                                       60
                      55
Leu Tyr Gly Asp Leu Pro Lys Asn Lys Glu Asn Ile Ile Tyr Leu Ala
                                      75
                   70
Asn His Gln Ser Thr Val Asp Trp Ile Val Ala Asp Ile Leu Ala Ile
                                  90
               85
 Arg Gln Asn Ala Leu Gly His Val Arg Tyr Val Leu Lys Glu Gly Leu
                              105
                                                 110
          100
 Lys Trp Leu Pro Leu Tyr Gly Cys Tyr Phe Ala Gln His Gly Gly Ile
                           120
 Tyr Val Lys Arg Ser Ala Lys Phe Asn Glu Lys Glu Met Arg Asn Lys
                                         140
                       135
 Leu Gln Ser Tyr Val Asp Ala Gly Thr Pro Met Tyr Leu Val Ile Phe
                                     155
                150
 Pro Glu Gly Thr Arg Tyr Asn Pro Glu Gln Thr Lys Val Leu Ser Ala
                                  170
             165
 Ser Gln Ala Phe Ala Ala Gln Arg Gly
 <210> 137
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..19
 <223> amplification oligonucleotide PG1ASe13
 <400> 137
```

	pattagata		19
accggg <210>	ggtcc agttgactg		
<211>			
<211>			
	Homo Sapiens		
<220>	nomo bupzono		
	misc_binding		
<222>			
<223>	amplification oligonucleoti	de PG1ASe14	
<400>			
cagaat	tccag catggag		17
<210>	139		
<211>	16		
<212>	DNA		
<213>	Homo Sapiens		
<220>	i		
<221>	misc_binding		
<222>	116		
<223>	amplification oligonucleoti	de PG1ASe15	
<400>	139		
ccgggg	gtcca ggcctt		16
<210>	140		
<211>	16		
<212>	DNA		
<213>	Homo Sapiens		
<220>			
	misc_binding		
<222>	. 116	d- PC1 A C-1 6	
	amplification oligonucleoti	de PGIASEIO	
<400>			16
	tccag gccttg		
<210><211>			
<211>			
	- DNA - Homo Sapiens		
<220>	——————————————————————————————————————		
	misc_binding		
	> 121		
<223>	amplification oligonucleoti	de PG1ASe17	
<400>			
	gggtcc agaatttctc t		21
<210>			
<211>	> 19		
<212>	> DNA		
<213>	> Homo Sapiens		
<220>			
	<pre>&gt; misc_binding</pre>		
<222>	> 119	1. pg11.g-10	
	> amplification oligonucleot	de PGIASels	
	> 142		19
	gtccag gatgcttat		1.7
	> 143		
<211>			
	> DNA	•	
<213> <220>	> Homo Sapiens	1	
	<pre>&gt; misc_binding</pre>		
	> Mise_binding > 123		
<2222	> amplification oligonucleot	ide PG1ASe24	
	> 143		
	atcaaa gcacagcatg gag		23

```
<210> 144
<211> 28
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..28
<223> amplification oligonucleotide PG1ASe25
<400> 144
                                                                        28
caaatcatca aagcacagat gtatcttg
<210> 145
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..20
<223> amplification oligonucleotide PG1ASe26
<400> 145
                                                                        20
atcaaagcac aggccttgca
<210> 146
<211> 26
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..26
<223> amplification oligonucleotide PG1ASe27
<400> 146
                                                                        26
agcaaatcat caaagcacag aatttc
<210> 147
<211> 28
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..28
<223> amplification oligonucleotide PG1ASe28
                                                                         28
atcatcaaag cacaggatgc ttatagaa
<210> 148
<211> 31
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..31
<223> amplification oligonucleotide PG1ASe35
 <400> 148
                                                                         31
 gtgttacttt gctcagatgt atcttgtgat t
 <210> 149
 <211> 23
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..23
 <223> amplification oligonucleotide PG1ASe36
 <400> 149
                                                                         23
 tactttgctc aggccttgca gta
 <210> 150
```

211> 2/		
212> DNA		
<213> Homo S	Sapiens	
<220>		
<221> misc_b	pinding	
<222> 127		
-2222 127	fication oligonucleotide PG1ASe37	
	incation originational reliable.	
<400> 150	- Land - and - Attacket	27
	tgctcagaa tttctct	. ,
<210> 151		
<211> 29		
<212> DNA		
<213> Homo S	Sapiens	
<220>	-	
<221> misc_k	oinding	
<222> 129		
-222> 125	fication oligonucleotide PG1ASe38	
	reaction original collection for the collection of the collection	
<400> 151		29
	tgctcaggat gcttataga	ر ہ
<210> 152		
<211> 20		
<212> DNA		
<213> Homo S	Sapiens	
<220>	-	
<221> misc_1	hinding	
<222> 120	212	
-1222 I20	fication oligonucleotide PG1ASe46	
	rication originatelectrae remisers	
<400> 152		20
caggaactcc	ageettgeag	
<210> 153		
<211> 23		
<212> DNA		
<213> Homo	Sapiens	
<220>		
<221> misc_	binding	
<222> 123		
<223> ampli	fication oligonucleotide PG1ASe47	
<400> 153	110ucton 011gonue	
	and the state of t	23
	aaatttctct gca	
<210> 154		
<211> 25		
<212> DNA		
<213> Homo	Sapiens	
<220>		
<221> misc_	binding	
<222> 125		
<223> ampli	ification oligonucleotide PG1ASe48	
<400> 154		
	acceptaget ataga	25
	ccagatgctt ataga	
<210> 155		
<211> 22		
<212> DNA		
<213> Homo	Sapiens	
<220>	•	
<221> misc	binding	
<222> 12		
<223> ampl	ification oligonucleotide PG1ASe57	
<400> 155	**************************************	
	tanathtata ta	22
	tgaatttctc tg	
<210> 156		
<211> 22		

<212> DNA	
<213> Homo Sapiens	
<220>	
<221> misc_binding	
<222> 122	
<223> amplification oligonucleotide PG1ASe58	
<400> 156	23
gcccaacgtg gatgcttata ga	23
<210> 157	
<211> 23	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> misc_binding	
<222> 123 <223> amplification oligonucleotide PG1ASe68	
<400> 157 cgaccatgac gggatgctta tag	23
<210> 158	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> misc_binding	
<222> 119	
<223> amplification oligonucleotide PG1ASe1X	
<400> 158	
ccggggtcca gagattgga	19
<210> 159	
<211> 26	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> misc_binding	
<222> 126	
<223> amplification oligonucleotide PG1ASeX2	
<400> 159	26
aaagtggaag gccctcttta acaata	
<210> 160	
<211> 25 <212> DNA	
<213> Homo Sapiens	
<220>	
<221> misc_binding	
<222> 125	
<223> amplification oligonucleotide PG1Ae1b3	
<400> 160	
gccctcttta acattgactg gattg	25
<210> 161	
<211> 24	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> misc_binding .	
<222> 124	
<223> amplification oligonucleotide PG1Ae1b4	
<400> 161	24
gccctcttta acacatggag gaat	
<210> 162 <211> 28	
<211> 20 <212> DNA	

WO 99/32644 PCT/IB98/02133

<213> Homo Sapiens <220>	
<221> misc_binding	
<pre>&lt;222&gt; 128 &lt;223&gt; amplification oligonucleotide PG1Ae1b5</pre>	
<400> 162 ggccctcttt aacaatgtat cttgtgat	28
<210> 163	
<211> 25 <212> DNA	
<213> Homo Sapiens	
<220> <221> misc_binding	
<222> 125	
<223> amplification oligonucleotide PG1Ae1b6 <400> 163	
gccctcttta acagccttgc agtat	25
<210> 164 <211> 25	
<212> DNA	
<213> Homo Sapiens	
<220> <221> misc_binding	
<222> 125 <223> amplification oligonucleotide PG1Ae1b7	
<2235 amplification origonacteoriae rounds. <400> 164	٥.
ggccctcttt aacaaatttc tctgc	25
<210> 165 <211> 28	
<212> DNA	
<213> Homo Sapiens <220>	
<221> misc_binding	
<222> 128 <223> amplification oligonucleotide PG1Ae1b8	
<400> 165	28
gaaggccctc tttaacagat gcttatag <210> 166	20
<211> 26	
<212> DNA <213> Homo Sapiens	
<220>	
<pre>&lt;221&gt; misc_binding &lt;222&gt; 126</pre>	
<222> 120 <223> amplification oligonucleotide PG1Ae3b4	
<400> 166 atgctggatt atagcatgga ggaatc	26
<210> 167	
<211> 31	
<212> DNA <213> Homo Sapiens	
<220>	
<221> misc_binding	
<223> amplification oligonucleotide PG1Ae3b5	
<400> 167 caaaatgctg gattatagat gtatcttgtg a	31
<210> 168	
<211> 23 <212> DNA	
<213> Homo Sapiens	

```
<220>
<221> misc_binding
<222> 1..23
<223> amplification oligonucleotide PG1Ae3b6
<400> 168
                                                                         23
tgctggatta taggccttgc agt
<210> 169
<211> 28
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..28
<223> amplification oligonucleotide PG1Ae3b7
<400> 169
                                                                         28
tgctggatta tagaatttct ctgcaaag
<210> 170
<211> 30
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..30
<223> amplification oligonucleotide PG1Ae3b8
<400> 170
                                                                         30
ccaaaatgct ggattatagg atgcttatag
<210> 171
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..21
<223> amplification oligonucleotide PG1Ae5b6
<400> 171
                                                                         21
tatctttgca tggcagcctt g
<210> 172
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..23
<223> amplification oligonucleotide PG1Ae5b7
<400> 172
                                                                          23
ctttgcatgg caaatttctc tgc
<210> 173
<211> 27
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..27
 <223> amplification oligonucleotide PG1Ae5b8
 <400> 173
                                                                          27
 ttatctttgc atggcagatg cttatag
 <210> 174
 <211> 20
 <212> DNA
 <213> Homo Sapiens
 <220>
```

```
<221> misc_binding
<222> 1..20
<223> amplification oligonucleotide PG1Ae56b
<400> 174
                                                                        20
ctgcccaacg tgggaaagac
<210> 175
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..21
<223> amplification oligonucleotide PG1Ae46b
<400> 175
                                                                         21
gcaggaactc caggaaagac g
<210> 176
<211> 25
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..25
<223> amplification oligonucleotide PG1Ae36b
<400> 176
                                                                         25
tgttactttg ctcagggaaa gacga
<210> 177
<211> 22
<212> DNA
<213> Homo Sapiens
<220>
<221> misc_binding
<222> 1..22
<223> amplification oligonucleotide PG1Ae26b
 <400> 177
                                                                         22
 atcaaagcac agggaaagac ga
 <210> 178
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> misc_binding
 <222> 1..19
 <223> amplification oligonucleotide PG1Ae16b
 <400> 178
                                                                          19
 ccggggtcca gggaaagac
 <210> 179
 <211> 56520
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> exon
 <222> 2001..2216
 <223> exon1
  <221> exon
 <222> 18196..18265
  <223> exon2
 <221> exon
<222> 23716..23831
  <223> exon3
  <221> exon
  <222> 25570..25659
```

```
<223> exon4
<221> exon
<222> 34668..34758
<223> exon5
<221> exon
<222> 40685..40843
<223> exon6
<221> exon
<222> 48067..48190
<223> exon7
<221> exon
<222> 50179..54519
<223> exon8
<221> polyA_signal
<222> 54493..54498
<223> AATAAA
<221> primer_bind
<222> 1991..2008
<223> upstream amplification primer 5-63
<221> primer_bind
<222> 2505..2525
<223> downstream amplification primer 5-63 , complement
<221> primer_bind
<222> 4091..4111
<223> downstream amplification primer 99-622
<221> primer_bind
<222> 4528..4546
<223> upstream amplification primer 99-622 , complement
<221> primer_bind
<222> 5475..5495
<223> downstream amplification primer 99-621
<221> primer_bind
<222> 5927..5947
<223> upstream amplification primer 99-621 , complement
<221> primer_bind
 <222> 8127..8144
 <223> downstream amplification primer 99-619
 <221> primer_bind
 <222> 8560..8578
 <223> upstream amplification primer 99-619 , complement
 <221> primer_bind
 <222> 11622..11639
 <223> upstream amplification primer 4-76
 <221> primer_bind
 <222> 12018..12037
 <223> downstream amplification primer 4-76 , complement
 <221> primer_bind
 <222> 11930..11947
 <223> upstream amplification primer 4-77
 <221> primer_bind
 <222> 12339..12358
 <223> downstream amplification primer 4-77 , complement
 <221> primer_bind
 <222> 12915..12932
 <223> upstream amplification primer 4-71
 <221> primer_bind
 <222> 13317..13334
 <223> downstream amplification primer 4-71 , complement
  <221> primer_bind
  <222> 13216..13233
  <223> upstream amplification primer 4-72
```

```
<221> primer_bind
<222> 13617..13636
<223> downstream amplification primer 4-72 , complement
<221> primer_bind
<222> 13547..13564
<223> upstream amplification primer 4-73
<221> primer_bind
<222> 13962..13981
<223> downstream amplification primer 4-73 , complement
<221> primer_bind
<222> 15994..16011
<223> downstream amplification primer 99-610
<221> primer_bind
<222> 16463..16480
<223> upstream amplification primer 99-610 , complement
<221> primer_bind
<222> 17304..17324
<223> downstream amplification primer 99-609
<221> primer_bind
<222> 17814..17832
<223> upstream amplification primer 99-609 , complement
<221> primer_bind
<222> 18008..18027
<223> upstream amplification primer 4-90
<221> primer_bind
<222> 18423..18442
<223> downstream amplification primer 4-90 , complement
<221> primer_bind
<222> 18699..18716
<223> downstream amplification primer 99-607
<221> primer_bind
<222> 19164..19182
<223> upstream amplification primer 99-607 , complement
<221> primer_bind
<222> 22589..22609
<223> downstream amplification primer 99-602
<221> primer_bind
<222> 23111..23129
<223> upstream amplification primer 99-602 , complement
 <221> primer_bind
 <222> 25098..25118
 <223> downstream amplification primer 99-600
 <221> primer_bind
 <222> 25657..25674
 <223> upstream amplification primer 99-600 , complement
 <221> primer_bind
 <222> 26537..26557
 <223> downstream amplification primer 99-598
 <221> primer_bind
 <222> 27022..27040
 <223> upstream amplification primer 99-598 , complement
 <221> primer_bind
 <222> 32262..32281
 <223> downstream amplification primer 99-592
 <221> primer_bind
 <222> 32823..32841
 <223> upstream amplification primer 99-592 , complement
 <221> primer_bind
 <222> 34215..34233
 <223> upstream amplification primer 99-217
 <221> primer_bind
```

```
<222> 34624..34644
<223> downstream amplification primer 99-217 , complement
<221> primer_bind
<222> 34473..34491
<223> upstream amplification primer 5-47
<221> primer_bind
<222> 34916..34936
<223> downstream amplification primer 5-47 , complement
<221> primer_bind
<222> 34702..34722
<223> downstream amplification primer 99-589
<221> primer_bind
<222> 35182..35200
<223> upstream amplification primer 99-589 , complement
<221> primer_bind
<222> 39591..39611
<223> upstream amplification primer 99-12899
<221> primer_bind
<222> 39971..39991
<223> downstream amplification primer 99-12899 , complement
<221> primer_bind
<222> 40531..40549
<223> upstream amplification primer 4-12
<221> primer_bind
<222> 40932..40950
<223> downstream amplification primer 4-12 , complement
<221> primer_bind
<222> 40629..40649
<223> downstream amplification primer 99-582
<221> primer_bind
<222> 41058..41078
<223> upstream amplification primer 99-582 , complement
<221> primer_bind
<222> 45729..45746
<223> downstream amplification primer 99-576
<221> primer_bind
<222> 46186..46203
<223> upstream amplification primer 99-576 , complement
<221> primer_bind
<222> 47879..47896
<223> upstream amplification primer 4-13
<221> primer_bind
<222> 48217..48236
<223> downstream amplification primer 4-13 , complement
<221> primer_bind
<222> 48902..48922
<223> upstream amplification primer 99-12903
<221> primer_bind
<222> 49331..49351
<223> downstream amplification primer 99-12903 , complement
<221> primer_bind
<222> 49830..49848
<223> upstream amplification primer 5-56
<221> primer_bind
<222> 50271..50290
<223> downstream amplification primer 5-56 , complement
<221> primer_bind
<222> 50172..50189
<223> upstream amplification primer 4-61
<221> primer_bind
 <222> 50573..50591
```

```
<223> downstream amplification primer 4-61 , complement
<221> primer_bind
<222> 50541..50560
<223> upstream amplification primer 4-62
<221> primer_bind
<222> 50940..50959
<223> downstream amplification primer 4-62 , complement
<221> primer_bind
<222> 50555..50572
<223> upstream amplification primer 4-63
<221> primer_bind
<222> 50964..50983
<223> downstream amplification primer 4-63 , complement
<221> primer_bind
<222> 50774..50792
<223> upstream amplification primer 4-64
<221> primer_bind
<222> 51183..51202
<223> downstream amplification primer 4-64 , complement
<221> primer_bind
<222> 51146..51165
<223> upstream amplification primer 4-65
<221> primer_bind
<222> 51479..51496
<223> downstream amplification primer 4-65 , complement
<221> primer_bind
<222> 51593..51610
<223> upstream amplification primer 4-67
<221> primer_bind
<222> 29734..29744
<223> upstream amplification primer 4-67 , complement
<221> primer_bind
<222> 51167..51185
<223> upstream amplification primer 5-50
<221> primer_bind
<222> 51667..51687
<223> downstream amplification primer 5-50 , complement
<221> primer_bind
<222> 51387..51403
<223> upstream amplification primer 5-71
<221> primer_bind
<222> 51826..51843
<223> downstream amplification primer 5-71 , complement
<221> primer_bind
<222> 51772..51789
<223> upstream amplification primer 5-30
<221> primer_bind
<222> 52199..52217
<223> downstream amplification primer 5-30 , complement
<221> primer_bind
 <222> 51850..51867
<223> upstream amplification primer 5-58
 <221> primer_bind
 <222> 52382..52400
 <223> downstream amplification primer 5-58 , complement
 <221> primer_bind
 <222> 52507..52527
 <223> upstream amplification primer 5-53
 <221> primer_bind
 <222> 52997..53017
 <223> downstream amplification primer 5-53 , complement
```

```
<221> primer_bind
<222> 52703..52721
<223> upstream amplification primer 5-60
<221> primer_bind
<222> 53142..53162
<223> downstream amplification primer 5-60 , complement
<221> primer_bind
<222> 53001..53018
<223> upstream amplification primer 5-68
<221> primer_bind
<222> 53521..53538
<223> downstream amplification primer 5-68 , complement
<221> primer_bind
<222> 53459..53476
<223> upstream amplification primer 5-66
<221> primer_bind
<222> 53920..53940
<223> downstream amplification primer 5-66 , complement
<221> primer_bind
<222> 54202..54220
<223> upstream amplification primer 5-62
<221> primer_bind
<222> 54681..54701
<223> downstream amplification primer 5-62 , complement
<400> 179
gtggatctgt gactgttcgc aggaagagag gagcgggagc aggacagaca ataactgata
                                                                       60
gtcaggagct gggtttggag ataaagaggg aacaagagaa agttaagttc tgtgttttca
                                                                      120
                                                                      180
tggcaaacat tgcacaaaag tttacaactt cgtgactaac agtaatctgg ggtgattcac
aacaaattta cacataaaca catatttact gactttatac acagcaatcc taacgtgaac
                                                                      240
acagaacctg ctttatcttt tcgcacactg ttctagtgta gagatgtctg gtctcagtta
                                                                      300
aagaaagcat aaggagcatt agttgtgcac actgtccaca cccgtgactt ttttccacca
                                                                      360
gtactaaacc tagtgcttct tacagtacag ggcaatgaca gccacagaaa gagagaagct
                                                                      420
cettttactg tgtaatgett cetgetggee ttcaaatact tgttacttga gagateteca
                                                                      480
ttcacctggc tttgtcccca aaggtcatca tctaccaatg atgttgttat ttgatgttaa
                                                                      540
                                                                      600
tcatgtataa agaaagtagc taccatcctg gccctgatta gaacttccca ctgaaatacc
gtcctgccta aaggtagcac aggtttccat tatggtggtg gtggggaggg ggcgggaata
                                                                      660
tatatatata tatatata tatatatatg gtaaagcatt cggcattctt ttaaagtaca
                                                                      720
actatecttg aaaagggtta catattaaac catttttacc acagecaaag gggaggagaa
                                                                      780
agatecaaaa gteetgtgga tetgetttaa cateaataaa acagttatee accettegta
                                                                      840
gettttagtg aaggetacaa aagtatgett tttatggatt acacatgtge acgeaactae
                                                                      900
 tttaattact acagaaaaaa acgaggctcc ttattaaaaa aaaatcagaa acaagtccaa
                                                                      960
 cagactetga ggaaatgaag caagagtgaa ttetgaaaag gtetaataaa cagtatggaa
                                                                     1020
 atatecttgt gggattgtte tteagetatg cataaacatg taattateat cattactgtg
                                                                      1080
 atggggaaaa acacggaccc taattctgaa acaccctggt agcgagagac gggcaggagg
                                                                      1140
 ggctgctgcg cactcagagc ggaggctgag gaggcggcgt ccccttgcaa aggactggca
                                                                      1200
 gtgagcagat ggggacactc gagctgcccc gcgacctggg ccgagctgcc tacaacctgg
                                                                      1260
 geceaggtge etgeaagaat tagaeeteeg ataaegttaa eacceaettt eteaetgete
                                                                      1320
 taattgtgtg catcccggcg cccaggggct tgtgagcagc aggtgcgcgt tccaggcagc
                                                                      1380
 tccagcgacc ctitaaacctg accgcgcgca cgtccggccc gagggagcag aacaagaggc
                                                                      1440
 acceggacce tecteeggee ageacceace tteacceagt teegteagte gecaccacet
                                                                      1500
 cccttcccgc gtccgcagcc ggcccagctg gggagcatgc gcagtggccg gagccgggtt
                                                                      1560
 gecegegeca cageaggtag etgtaetgea aetgteggee caaaccaace aatcaagaga
 cgtgttattg ccgccgaggt ggaactatgg caacgggcga ccaatcagaa ggcgcgttgt
                                                                      1680
 tgccgcggag ccccctgccc cggcaggggg atgtggcgat gggtgagggt catggggtgt
                                                                      1740
 gagcatccct gagccatcga tccgggaggg ccgcgggttc ccttgctttg ccgccgggag
                                                                      1800
 eggegeacge ageceegeac tegectacee ggeeceggge ggeggegegg cecatgegge
                                                                      1860
                                                                      1920
 tgggggggga ggctgggagc gggtggcggg cgcgggcc cgggcccggg cggtgattgg
 ccgcctgctg gccgcgactg aggcccggga ggcggggggg gagcgcaggc ggagctcgct
                                                                      1980
 geogeogage tgagaagatg etgetgteee tggtgeteea caegtactee atgegetace
                                                                      2040
 tgctgcccag cgtcgtgctc ctgggcacgg cgcccaccta cgtgttggcc tggggggtct
                                                                      2100
 ggeggetget eteegeette etgecegeee gettetacea agegetggae gaeeggetgt
                                                                      2160
```

actgcgtcta ccagagcatg gtgctcttct tcttcgagaa ttacaccggg gtccaggtga 2220 2280 geogeeteec getecegggt eteggegtee accegagete eegggggege ggacetetee 2340 2400 gtgccgcctc cccgccttcc tctccgcatg cttcctgccg ttctgccgag atcgctctct aggaagetgt ggetgegteg teetgagget aegagtggga eeegeegeee ettteeeege 2460 ccctcgcctg ggtctgatgc tgcttagcaa agtgggtgca gatgcacgtt ttaaataata 2520 gggcacgcgt ttagcagttt ctggcctttg gtccaaagag gtggtcatgt tggaacagat 2580 cggagacgtc tacactccga agtgcgcttt tacagtgacc tcttgaaaca gaagtacaat 2640 2700 toggtottgt gttotttoco otggacaagt gaaagotggg ogaagaaatg aatacatttg 2760 ttaaccgtag aagcctaact agatacaatt cttgccaact ttaactgggc ttgaatgtgt gggtgatctg ttgtctgatt actttctttc tgttactgtt tctctgtaga gattggattc 2820 gtagattaaa cttgagaaac aaaccataaa agtggaaggc cctctttaac agtaggtatt 2880 tgaagtgtta taaaaaaaaa aaaggtgaat ttttctttta tttctcagtt tgaaagaaca 2940 3000 getttattet tggttattee taatgteeae etagteetet titaetttie ttggtagggt 3060 tagggtggca tggggaaatg ggacggtatc attttgtctt tttaactttt ttttttcca cctacagcag ctgtttttac cctgtggtca gtcaggtact atatttagtt tgcagttgca 3120 ctgctgatcg acccttgatg gccccagttg gaagttgttt ggggggaagg aactaggaga 3180 3240 ggccagggcc tccatttaaa ccagtgtctg taagtgtctc cttggaagga aaaaaagata 3300 ctgttccagg tcatggtttc ctggtagttg acgtttaaaa tgggcctcat ttaaaaaattt caataattca ggctaatttt ttccctttat atggtaactc caccaagttt gtctaaatgt 3360 atgattttta tcatgattaa gtttttactt ccacatcatg tgacaactgg cctgggatgg 3420 gatataagct cagaacacaa agtcattcac ctgttaaaaa aataattcta tctgtggcgg 3480 3540 gttatgttat ttttgttcaa agaggacaca atatgatgca gaatacacca ttgaaggatt 3600 ttttggtttg gcaagttctt attttttaa atggctgtaa aacctagcag tgtttctgaa attgcatacc ttacctgatg ttcagagatc cgatttactt cttgatttcc cagcaagtga 3660 ttttgaaaac atttaatcta atcattcccc ccaccgtctg ttcaaatcaa aggaagtggc 3720 atccagcact aattttcatg catttatgaa aggatgcctg aggaccctta agtataattc 3780 aaaattttgt ttaatgtgtg ttccttgatg aagttcttta ggagtcgtag aacgaactga 3840 ttgcccactg atcatcaaat gcaagttatg aacatttaat aaaaatttaa aaccaagagt 3900 ttcttgttcc tgcattttta tttttattgt atggagggga caaataatta ttttctgttt 3960 agtaacagag cagggtattt tgaatttatt agggtctttt tctgcagtct gggtttcctg 4020 tgtacacaaa gctacctttc aatatttttt attgtttctg ttaagattaa atcaatagag 4080 gaataaatag ctatcttcaa acataagacc caaaggaaaa agatttatag tgatgttctg 4140 4200 tcaccttatt ttttacctgt gactttgtac cattaacttt gtcactgaga tgttttgatt 4260 aaaattttta gcttgctttt cttgttttgt taggacactc tttttttctt gaattgtttt tatcagcttt cgtttgcaag gctagtgatg attctcttgt tctgtataaa gtattgttga 4320 ctcatttctg aagggagttt tagtaattta agaggttata agtttttaaa taaaaggttt 4380 4440 4500 acacctttca actctaggtt taaaaaataa gtggttcaca gtagttcttg cagaagaata ttttctttta catagaattt ttaagctgaa gagaagtagt agtaggtcca tgagatttat 4560 gatctgtgct tggcaggtaa acctgcttcc aacaaattta gttggatttt tcttggattc 4620 tgggtaaata cctttttctt ccccagtttc actactttat tttcatatgt atctctgaga 4680 4740 tagagaaata tttcagtcag tgctgctaaa attgttcctt ataactcgtt tatcctttta ggtccttcca gaatctctca ttggtactga aactcaaatg ggtactttct tcaccattta 4800 4860 tttctttaga ataagtaata agaattttat aagctttttt atatttcacg taatttgaga ctattgaaaa tccagttaag tctctctact gtgttgagag gcattgattc aagtacctgt 4920 gttactttcc tgtgctgcca aaacagatca cctcaaacta agcggcttaa aataatagaa 4980 cttaagttct cgtgattctg gaggccagca ctttgaaatc aaggtgtagg ctcaatttta 5040 ctccctctgg aggccctagg gggaatctgt tcttgtgggt ttcaacttct ggtgactggt 5100 ggcattcctt ggcttggggc cccatcactt caacctctgc cttacagtcc ttgctgccac 5160 ctcttctgtc tcacatctca ctctcccttt ctcttagaag gatgcttgtc attgggttta 5220 gageceacet ggatatteeg ggatgatete tteateteaa gateettaat tataaetgea 5280 aagagccttt ttccaaataa gaaaacattc acaggttcca gggcttagga tgtggacaca 5340 ttttttgagg ggctgccctt cattccccca caacaatgaa ctccatagtt ctgcctattc 5400 agtattttgt agttatttcg tagtttaact tgccttattt ctttaggtat ttacgtatta 5460 aagcattttg gtctctgctt tctttaacag agaacctggt tttctgtaat aagtttactt 5520 actttcccat aatcttttag tttcttattt acagatttac cttcacatat cccttaagta 5580 gaacatttga ttaactgttt tattttcgga acaaatctgc attctgtata ataaccaact 5640 5700 tattcatatt tcggtattct tttaattctt atctgattct gaaattacca tcttgtgatt atatatata atatatggaa ataactgaaa tottgataaa ttaaaggtga tataacttot 5760 aagacaatta attatgtatg atgtggtgaa tatactggtg tttggtttgt ttgccactta 5820

aaagccctat	ctataggata	ggaagtaact	tgaatgtgga	atgcttagag	actcagagta	5880
agagggggta	tatatatcct	tgagctggag	tttaaggaaa	acttatggga	aattaaaagg	5940
agaggeegea	tactgacaga	ggattgcgta	ggactcatga	aaaaggaatg	aagttacctt	6000
aaagttggag	atcgtgagtt	aacgtgaaac	tagatttatg	ttagtttata	gcctagaatt	6060
ctatectace	aatctagata	tatcctaaat	gttgagatag	ctgcataaac	aataactgta	6120
atcattataa	taaataatga	caaatctttt	tagcatgttt	tgtgaagctg	ataaatgtta	6180
attercates	ttcaaatgtc	agaattettt	tttctttact	tcttttttaa	aaaatttctt	6240
ttc/cccatt	cctatgcaat	acactgaaaa	ctgatcattg	aaatttgtag	gccaaaaaat	6300
tantonage	gtaatagatt	agaatttaga	ttttttgag	tcagggtctt	cttctgtcac	6360
caaccaacac	gtgcggtggc	accatcatoo	ctcattgcag	ccttgaatgc	ctgggttcaa	6420
atastactac	ggagtagctg	ccgtgccatt	atttctagct	aatttttaaa	agtttttgta	6480
gegatecete	ctttctgtgt	tacccagact	ggtcttgaat	tcctggcctc	aggtgatcct	6540
tetacettaa	cctcccaaag	tactaggatt	acaggtgtga	gccaccatgc	ctagccccta	6600
ateastatto	taattaccga	tttatcttgc	ttaaatcagt	tggtaacact	tggaatttac	6660
ttcacaatat	attttacatt	agtggctctg	actoctaatt	ccccttctc	caaatgctaa	6720
tetagaatat	caataaaatg	cacagttett	aagtttatat	aaaataaaca	ggttttcagt	6780
tocatacaa	taagtgtaaa	atactotoaa	aaacacaaga	aagaagataa	agaatttaag	6840
attttaaaat	ttctctaata	tocccttaac	ttctccaagg	attcatactt	ttttttgtaa	6900
accepacac	cacactgttg	cccaaaccad	aggtgcagtg	gtgcagtctc	cactcactgc	6960
gacagaacct	cccgggctca	aggggtgctc	ccacctcage	ctcctgagta	gctgggacta	7020
aacctetgee	gcaccatgcc	cagctaattt	tttttttaa	tattttttag	tagagataga	7080
caggtacaca	tgccatattg	cagecaatetaa	ttttgaggtg	ctgggctcaa	gtgatccgtc	7140
gacgagattt	catgcttagc	trattratar	tettaactga	aacattotto	caagtttctc	7200
ettgateeae	aaggcttttt	atctacacac	catttataac	togatette	tttgtgtagc	7260
agaaacagtc	caaactaatc	ctagagaa	aatgagttaa	atttatattc	tgaatcttgc	7320
actgattcat	gccattcatt	agaatgaaac	atattactt	agaattggag	aaggagett	7380
tgtaaaagca	tagtctactc	agaatyaaac	cacttctaca	ttctttctac	acttctgcca	7440
ataagtcatc	cagcgtcgtc	tatastacat	atactcctaa	caagaatatg	aatcatacct	7500
aaatgttgcc	attttactct	tototactta	tttaccattc	atgtgaagac	cttaaataga	7560
tgtatcctta	cttccttcac	tttagetgag	agtgacagga	ctgtgtaggt	ataaatatat	7620
tettaaattg	gcttatttaa	ggaggataat	aaaaactttt	actataggaa	attaaacatt	7680
tectgeatte	atacaattcc	actctaacac	aattaaatto	tagttaggga	actocttaac	7740
tcccaatcaa	tataggaaaa	tactasassa	atgraactag	aactctattt	ttacacttta	7800
tractagact	cctctgtgaa	caecaaaaaa	atttcaggtt	gcatttgtgt	atagttttt	7860
taaatataaa	ttttctattt	taaaatcaca	gatgcaatta	tacattcaaa	cactgccaca	7920
aatgeetgat	aaagttaaag	tttcccctac	tectacaeta	cgtacacctt	tcctaggtac	7980
atactttgag	ggtgtgtaac	tttagatttc	ttccaagagc	ttttgagtaa	gtgtttgaat	8040
ateceagett	ttctttagtt	aaatgaactt	cttacagasc	agtttttag	tacagtagca	8100
tgtgggaagg	ctgcatacct	atagaatac	ctctgtgcca	ttacgatgga	aggcacggga	8160
cgaaatatac	ccgtatatac	ctactttact	trecetett	totatattto	tctgattttg	8220
aaacagcact	gcttctcaag	tagattaga	acttaacttt	teetttaeta	ttttctcatt	8280
tggagetgat	. gcttctcaag	gaattaata	ttagtggttg	gaccattcaa	tagtaagtaa	8340
tractatygu	. ctcctaacta	totagaacco	agatecetta	attectgeaa	tattcccgtg	8400
tgactitte	t gcaagggacc	atcatagate	ccaagggata	caatgtttga	tagacaatgt	8460
cgcacactg	. cccaggcgcc	cattctagto	taggaagaca	aacaataaca	agaaaatgaa	8520
accigccati	. alggaggica	attatttaa	ctaataaata	agaggtagg	gtttggaaaa	8580
aatttactgt	gccatgccag	ttatataaa	ctctctaaaa	gggccagctt	ggaaggtaat	8640
ccttactgag	, caaytyacat	. tratrottac	, ccccgtaaa	agagtggatg	ttaagattga	8700
gragicated	aggigagaaa	tagtggttag	tgatagggct	ttgtgattg	a rtgtggagga	8760
aaagaattt	addictation	taacacacto	agtcgcagtt	agtgagtgct	gctgtgtgca	8820
aaaayaaya	g ggtgggttag	ataattccat	ctttacaaac	taggcaccat	tetteetett	8880
agtattgtt	callalgeas	acacccata	ttcacatctc	tagtagccta	a gccaggagtt	8940
tacagaca	a gyaaaagggc	gatgetetg	ctggcaatgt	ggttatatt	g gttgaaatga	9000
ccaggcact	t atticiting ac	ttcatctag	r aaagacatga	actoccaat	acaatatagg	9060
gacccccta	a aattamama	atatttatt	actttoccat	acagaggta	a agtaactctt	9120
tassetass	a aarrayaya t otttoottoo	, gegeeeace	aaggctataa	a aaattactto	g gagtttttac	9180
tadagtade	a catasttes	, gulaguyya;	t tagggaaaaa	aggttttca	a ttgataacat	9240
natanasat	y tytaattaat	accatacca	t tcaaggttt	t ctaaattct	g ccccggttaa	9300
aataaacat	y ayyayılıyı t cattaattt	attetagte	t agetttteet	t tctgaacca	c ccctcccac	9360
atterace	c tectetete	r aattocaac	t caagecette	g cttttctcc	a tctgtcatga	9420
tattacco	a totoattot	agggtaact	t ttatqtaat	a ttaacatat	a taatactgat	9480
Lyctacted	a coloque	- ~3330~~			_	

						0540
ataacattag	catattttaa	tgtatggatc	atctcctctg	caacattgta	acctcttgga	9540
datodcaata	atoggaagaa '	tgacttgatt	ttacttttc	ttttaacaaa	aatggtggag	9600
tagtctgggc	acaatataac	tcatgcctgt -	aatcccagca	ttttgggagg	ccaaggaggg	9660
togatcactt	gaggt.caggc	attcgagacc	agtctggcca	acattgtgaa	accoration	9720
taccaaaaaa	atacaaacac	ttactgggca	tggtggtgtg	tgcctgtagt	cetagetact	9780
candaddctd	aggtgggaga	atcacttgaa	catgggaggt	agaggeteca	gettgggega	9840
caggaggeeg	ccctgtctca	aaagaaaaaa	aaggtaaaag	ggccaggtgc	ggaggctcac	9900
actagtaata	caagcacttt	gggaggctga	ggcaatggat	cacctgaggt	cgggagttcg	9960
gerggraace	gaccaacatg	gagaaacccc	ttctctacta	aaaatacaaa	attagccggg	10020
agattagtt	cctgcctgta	atctaagcta	catgggaggc	tgaggcagga	gaatcacttg	10080
cgtggtggtg	acagaggttg	taataaacca	agatggcacc	attgcactcc	cgactgggca	10140
aacccaggag	acayaygtty	caaaacaaac	asacasacas	aacaaaacag	agagaaaagg	10200
acaagagcga	tagggaattc	tagaacaaca	ttctctccaa	atgtatatga	atctcacttt	10260
cagagtactc	agatttttga	atacastasc	tagttgataa	attractet	aacagggtac	10320
taagggatgg	agatttttga	acggcataac	cttaaatada	taaaaaaaa	agaaacatga	10380
ccaagtctag	tgagtccgat	catteric	cctaaacaga	2241234234	agaattetea	10440
ctccaccctc	aagagtaagg	cagaatgagc	adagteagag	tonnaccad	tactttttct	10500
cgcagccagc	agtgcagaga	aaccttggtt	tagttgtgaa	ttattatat	tatatataga	10560
aatttttgag	cctatgcaat	tctccaaggt	tttatgttgt	tecttetget	cttctgtagg	10620
caccagaaat	caaaacccca	aataagaaag	tgttacttga	agatettaga	gracticatic	10680
gtgtataagt	gtaagtgata	tttggaagac	gactttactg	egeteeteea	gettggeatg	10740
agaattccag	gggcggaaag	aaaggagggt	gatggtacct	ggaaaggaga	greatgread	10800
gtcccagcca	catattaagt	gctaaccacc	tactgttaaa	aggigiaatg	ttctagactg	10860
acaaaataca	tagtctctac	cgtaaagtaa	cacataattt	agcagtgcag	aaagatgtca	10920
cttaaaagaa	aacttgaata	tatgctgaga	tagttcacaa	attaaagaaa	tgaacaaaya	
actroaggaaa	taaaggagga	atacaactgt	gtccaaatga	atacttaact	gggrgggage	10980
tattacatat	graagcaggt	ggttcaccta	aaagttggat	gtaacgtagt	Laacyccayc	11040
tettaataca	crtacatatt	gcattgcttc	cgggcttaat	ttgtgttcat	atayyaataa	11100
attttttatt	ggtttttaat	tttactcctt	gtaattccgt	ggttgatatt	Caaagtyaaa	11160
aaaattacat	aagettetaa	tatatgagaa	gtcttctcac	ttgacatttt	LLattiggaa	11220
tttttgcaga	gagtagtttt	gtcacagtca	aaagattttg	ggatettgea	gegagaaace	11280
taggtgtaat	tectatttet	ctgccattcc	gtatgtcatc	tggattaagt	gicaactici	11340
carteteaar	attetegtee	ttaaatggaa	tactttttgt	catgctattt	tgaagacaaa	11400
atgagataat	acotoaaact	gcctagctca	gtgaatggta	catcatagat	acteagaaaa	11460
aacacaccct	ctaaaataaq	aacagtacca	aaagacagga	tgtaaaataa	gggcagtacc	11520
aaaagacaca	tacatactaa	gtgtatgaga	aagaactttg	tggccttctt	gggtggcaca	11580
gaccatagca	gttccacage	atgacgtggt	tgctgtgggt	ggtagagcag	acatgoogce	11640
ccccatcact	acctaacttt	gatgettget	ttcttcagct	gagaggacgc	agetgtgata	11700
traarrtctt	gratatacag	tcataacctc	acatttccaa	tttcctgctg	gcagaaccca	11760
cantictacas	catacaaaca	ccagagttga	cgtgagacag	acagcataca	gaggerigia	11820
acatecttet	ggaaaacact	gtgtaagctt	tcagtgcgaa	taaacatgat	cagiggcaag	11880
ttctattaga	tataatetae	aagcatcctg	attttactgg	gcaagactat	. gttgatttat	11940
aggeggetga	: roattccato	gatagcccac	tactagtatt	. ttcacaaatt	. tcacaagaca	12000
ttcttactgc	aagattgccc	tattettata	atactgctgc	ccttttagct	. teattigety	12060
ttcagactaa	a acttggagac	tacagtcagt	. cagagaactt	gctaggccac	ctctcaggtt	12120
attettteat	- tootgateat	cctcaaaatt	. ttgaaaaaga	l aattgtaaaa	actacatcag	12180
caacatatac	r acttatated	ttgagaagca	gcagttaatt	: acctaaacac	agcaagtacc	12240
ttagaagtgi	- atagaattaa	attocactat	gcaagggatc	: aagtaacaat	, aaaattatga	12300
ttagaatgat	t atcaagagga	attctgattt	: ataacaggct	: atgaatgagi	acctttccat	12360
antenaanai	r totaaaaatt	: tottttaagt	: gcaaacagtt	ttttattcag	g ctttgaaaat	12420
ggccsaaga	a aatctqqaqa	aagattatca	ggatttaata	a tggtgaatta	a tatggcatgt	12480
gaccigcac	t dattiggago	tttagaacat	cacatattct	tctgtttgg	a cagaccactt	12540
adacacccg	e agaattttt	rocacattat	tttacattac	gttcaaaati	t cctaatgcat	12600
agtagga	a ctraenttre	gttagttcag	tatggcaaa	aaaaqqcaa	a taaagacaga	12660
ggrgggaga	a cigaagiica	taadccatt	acatadaaa	taatagttt	g ggaagtagta	12720
ctacttyca	y yaktitiday	trassarst	, adaaacata	a agcetteca	t cacaattccc	12780
ggcaggaat	- caatatety	. cyaaaayac	a tretecation	c atacaagag	g gaacctgatt	
acceggaac	t totattant	atttaatea	a ttatutuut	t cacacttct	t ccaaatattt	12900
atgaccatc	t tergergge	- atteggtage	a detastade	a tcattctcc	a atcatcacta	
gcaaatcag	a catcaccati	tagataga	t ccagcttta	c cacccaaac	a agctaagaaa	
ttacaggac	a cecetggag	a thatttaca	t tttcaacaa	g atctggaag	a tcatattaat	
aactgttgg	a accadance	t totottasa:	a aatotooto	c taatggtgg	t attctacatg	
gaaacgitg	a cyclicatio	. cocceaaa				

						12000
ataatcgtgt	tctaatccga	gtgaacctga	cgaaaatgga	aggtttggag	tcaatgcaaa	13200
gggggatatg	atcagaagat	gtctgtgatc	gtgtcctgag	aagcaccagg	aacacctttg	13260
acctcagtga	ctctcgattg .	aagagaagac	caagttgtat	tgatcagtgg	ttgggacttt	13320
acagaacaca	cccatgattg	ggttgtcctg	ctttttaaag	ccaactgtga	gagacattct	13380
ggggaactca	tgcttctagt	tctacctatg	ctgcatatga	tgtagtggaa	gaagtgctag	13440
aaaatgagac	agacttccag	tacattctgg	agaaagcccc	actagatagt	gtccaccagg	13500
atgaccatgt	actataggag	tcagtgatcc	agctaaccga	gggcttatcg	ctggaacatt	13560
ctggacacaa	tttgatcaac	ttatcaaaaa	aaaacttgga	atgacaattt	ctggtgccag	13620
attaccttag	aacctttgca	aaaatagata	gagatagttt	tccttatgat	gttacatggc	13680
ttatttttaa	aggtaatgaa	aactacatca	gtgtaattcc	agcatcataa	gtcagaacag	13740
tgcttgtcaa	ggggcgttac	cacacacttg	aacagatttt	tggcagatga	cttgggaaca	13800
aggeteetee	atgtttgtaa	tgttgaccac	acaagttgaa	tgtggcagag	ttaaatgacc	13860
ccaatattgg	ccagaaccca	caggaagttc	atcctatgga	tgctaccaag	ccttctgcca	13920
ctgagaagaa	ggaagcactg	tctttatctt	caggaagatc	acactgctgt	ttaaccaaga	13980
gaaaaattag	agagtcatca	atcacgcaga	tccagtacag	agggtggcct	gaccatggag	14040
accctgatga	ttcagtgact	ttctggattt	tgtttttcat	atgcaaaata	agagggctag	14100
caaggaaaaa	ccccttgttg	tttcttgcag	tgctggagtt	ggaagaacca	gcgttcttaa	14160
tactatggaa	acagccatgt	gtctcattga	tctcattgaa	tgcagtcagc	cagtttattc	14220
actagacatg	gtaagaacaa	tgagagagca	gtgagccgtg	atggtccaaa	cacctagtca	14280
ttacagtttt	gcgtgtgaag	tactattttg	aaagcttatg	aagaaggctt	tgctgaagaa	14340
адсававада	aaaaaaqaac	tttgtcatct	gttaggttcc	atttattgca	tgataattgt	14400
atttatatta	attattgggc	aagtagctgt	ttgctatttt	gatcttattt	cagaagggca	14460
taataatttt	actattcaat	gaaacgtttt	aaacggggta	gaaaaagact	agtttttgta	14520
toctttacag	cagaaatctt	ataatgatta	actggtaata	tatttcgttg	gcataaaaat	14580
acatttaaaa	gttcaagtaa	ttataaacat	tgtaaattgt	atatgtaatc	atattgaaat	14640
tgaaattctt	tatagctgta	cttctgtgta	atcaaagact	ggggagagat	agactagcta	14700
actetttete	ttatccatta	atcacttaac	agagttttga	ataaaaagtt	ccatttcatg	14760
ggataagaat	aatgacaggt	taacctattt	tagttggtta	ctatgttcta	ggtgttgtat	14820
gaagtagttt	acatagtttc	actgatttca	ctacaatccc	aggaggagta	gttactatta	14880
ttacactcat	tttacaggca	aagaaatagg	tttggagggg	ttgggtgttt	tgcccaagtt	14940
ctcatcgtaa	aatgacagat	gaggattcaa	attcaagtct	taattgaagt	ccattacttt	15000
agaacctacc	tcttagtggc	tcttatgtta	cagtataagg	gagagcagac	tgttccttta	15060
cccttgtagg	gtagctaggg	cttgtgaatt	aagagactga	ttaacaggag	aagaggcata	15120
cacattttat	tgacgttagt	atttttacat	gcacagggaa	ggagggtttt	atttttattt	15180
rtatttttat	ctttatttta	aagagacagg	ggtcttgctg	tgttgccagg	gctggactca	15240
aactcctgaa	gccaagcgat	tcttctgctt	gagattcctg	agtagcaggg	actataggtg	15300
tactcctcta	tacttaacta	aagaaggggt	ttgtatgtga	tttttaacaa	aggctgataa	15360
attotoaaga	agtgactagt	caaaggagaa	gaggatttca	gctcccaggg	gtggtaaatt	15420
gtgggaagat	gactaggaaa	tgtatagtaa	taaggtttgc	tatgcaggtt	tattttgcca	15480
atttctaatc	tcctaataag	ggacagggaa	acacctttac	agatggaaat	tcatatcacc	15540
tttccacagg	gaaatttatg	tcctgcctta	ggcagttagg	ggaagggcag	agaattette	15600
ctgtatctgc	tgtgtctcag	gtgccttcag	ctcaaaataa	tccttatgcc	aaagtagcat	15660
atttgggtgt	ggcatattct	ctgatctctt	tcaacagcat	catctatact	taacaacagc	15720
aaaagtttt	. tttaaaaaat	catgtttcaa	gatttgcatg	tggaagacaa	atggacatga	15780 15840
ttgagataaa	tgaagaatat	atattttta	acaaagaatg	ctgtatattt	atgtetetgt	
gacattgtgt	tatggaggct	aaggtgttaa	gcatgtgatt	actttagatg	ccgtatgact	15900 15960
acctgttttt	. aagattaaaa	aagaatcaat	aggcagttta	tatgcatggg	agcaagttaa	16020
aaacaacaca	gatgtgatga	aggcgaggtg	aaactggtcc	gcatctaatt	eaggeettet	16080
cctgaaagcc	agtgtgtgca	agataaataa	gtttgtttga	cgaaagcaga	ataactagtt	16140
tgtcctttgt	gatgaagata	gttattcaga	aatcatttt	attggctacc	tetgaattaa	16200
taaatgaaaa	a gagaaatttt	tttttctgta	ggggatgtct	gatgagttet	taaaaagtyy	16260
atgaacctga	a aattatcatg	aacaagcaat	tataatgaac	ttaaaattac	teaaagagtt	16320
atgaaaaaca	a aaaagaaaag	ccgtatgttt	ccttgtgcct	cattttgaag	cyacaaatta	16320
tttgcagggt	acatttgtag	acggaactaa	ı çgtgatttaa	aaaatgagta	ctagatttac	16440
agaatgatgo	c ctttaaaaag	teactggtge	: actttaatta	tttattat	gtttattctg	16500
aaactacct	t tattttgaaa	atgaggtata	a getttgeeta	ctggtgacaa	aagtgtaaat	16560
aattcacta:	a acatototta	aaaaccagct	: tggtgctagg	r ctcttggggt	: agaaaactga	16620
tcaggccat	t gaggagctca	tagtccctaa	a ggggctgggg	actigicati	aggtgtgcag	16680
tgtgttctg	g atgctcctga	a aggagtgtgg	gcaggtgcgc	accaccatgo	ctggctaatc	16740
tttttataa	t tatgtagaga	a cagggtctgg	g ctgtgctgcc	: catyctyggt	ttgaacttct	
gggcttaag	a gatcttccct	ccctgcccct	t accgacccc	g cccgcccact	. deadeteage	10000

ctccccaaag	cactgggatt	gcaggcatgg (	gccactatgc	ctgggctgtg	caaaactttt	16860
aaatcagtgc	atactcaatq	gtcttgatgc -	aattctggct	tgttggtaag	agaatgggga	16920
tttactcaca	agccacgatg	tcacttttaa	ctctgaacag	atcaagctat	tggtattact	16980
catttatqtc	atcgataaac	tttatgaata .	aaaactcatt	gtgcaaatat	ttaaacatac	17040
tacatacata	gcactgtgca	gtttctaagg	aaagtaatgg	aaacctttgt	cacatecetg	17100
acttccagaa	ctttatotta	tctaaqtqca	tttgtctgca	aagttgttgg	gttaattgee	17160
cetttette	ttetetttt	aagatattaa	taaatagtgt	catgaccaaa	agataatcct	17220
tatocacaao	atagatctaa	aaagccttag	ctaatttata	atcttgcata	atccatgatg	17280
acaacatcca	gaaacaaaaa	tocccagaat	aaaaacttag	caccattagc	agccatttcc	17340
ttttaagtct	ttacaagtat	actcccagtt	tcttgaaaaa	tttattctaa	aatatgtaag	17400
acacacaaaa	cagcagaagg	actaatacag	gtacatcgaa	cacctgtgtg	cctaccgccc	17460
acacacaaaa	taaactggaa	tgatgtttct	ctcatactta	cagaataaag	ttttaatctt	17520
tagettagaa	ttcaaaagac	ttctgccatt	ccagttcaga	gccacccttc	tggtctcctt	17580
caycatggaa	cgcgacactg	cccatgtacc	caacaggcct	ccagggttac	tgcttccatt	17640
geteeteage	ctcatgaaca	ttttccttca	totcatctgc	cagaatccta	cctaataata	17700
cyclectatt	gcagtttaca	attattaaa	attaaaaaaq	attatatacc	ctttagtgtc	17760
etectgetet	aaaaaacaaa	tttaaaacct	taaaaaaaata	ccatattttc	atagtatttg	17820
ctgaaaaaag	cattacagtt	catatagaca	tatatatat	ttttactaga	ttgattgtgg	17880
cgttatgtct	ggaagatata	tettatgate	agtotttat	atattottao	caatcaatga	17940
gctctttgaa	ggaagatata	atanagatat	tgattattct	attttaattt	attaccotta	18000
atgcttgcta	tatttttctc	atgaggatat	ctaractatt	tatgatetet	gaggeteetg	18060
acctgtacta	tacataactg	etttetgtae	ctgagctatt	acaacatcat	tcatcaattt	18120
tgagaaatct	aatttttgtt	aatcatggat	ggaaatattt	atatcaaagt	aatottttt	18180
cttcacattg	tcttcctttg	tatattacag	arguittaaa	acaccadage	aaatataata	18240
tgttttatct	tttagatatt	gctatatgga	gattigecaa	addacadaga	taggtttttc	18300
tatttagcaa	atcatcaaag	cacaggtttg	tatttcattt	ttanatana	taggttttt	18360
tacagatggc	acatgggcat	tcaaaatacc	gttcttatat	tradatyday	attttacttt	18420
aaaacagcaa	ttttctgtgc	agatattaca	cetgttettg	tatttttgtg	acctatatta	18480
ttggaaagtc	agaaacttga	aagctatgaa	ttttcctaaa	cttaccttct	acatatttaa	18540
gatgtaagta	agctatcttc	ttacttgctt	gctttgttt	teettigigi	agetettaa	18600
agagtgtatt	cattctttt	gtaagtgatg	tttctagaag	tagcattggt	gggtcgaagt	18660
gtgtatacat	tttacatttt	tgattgctaa	gctgcagaaa	agctgtattg	gtatgtaagt	
actedtttcc	tractatoct	catcatttct	agtgtctgct	CTTCCTTTCC	ttetteaaat	18720
gggtttggtt	taattctagt	tgctactgtt	ccatcagagg	aattgcagag	aactggtctt	18780
cassacagto	cagtatatac	tttaggtgaa	gatacttcta	aaaacctttg	cactetgagg	18840
taattotaga	gtcccaagaa	tttgcaaaaa	gagtacattg	tcagcaatat	LLLLCCCaal	18900
ggtgacatct	taatataact	gtagcacagt	agcagaatca	ggaaattgtc	actgggtaag	18960
atacttttta	attotocaaa	taattcagcc	ctccaaaaaa	atcccacttc	Leatytttt	19020
assectated	, ctacttttga	tacatactta	ctaaattqca	tttttattac	tttaaaaaat	19080
ataataccta	. daadctcaaa	gctggaaaca	gcctgatcaa	tatagtactc	ctaayctaaa	19140
aacaacctga	tcaatatagt	actcttaggg	aaatcactta	rgeergrage	LLCLLCCAGG	19200
ttttcttcct	· atcadctate	tcttcatgat	tttgtggttt	ttattactgc	Clacaccaca	19260
ratrarratat	agaagtaaa	agaagttaaa	atgcattttt	ctcaatttag	tgaattaaty	19320
attacattca	a datttatagg	acaaqqqttg	aagctacaag	gggttgalay	gaattttgat	19380
~+ ~ + ~ + ~ ~ ~ ~ ~ *	- attttcccca	actttattac	atgactggtt	cagactatt	lattiaatta	19440
catttcacto	ttoocagaaa	tagcaaaaca	gtcaaccaat	ggtcaatgct	gergagaact	19500
ctaacctata	r capacatatt	gactattta	Cilcidatac	Catterger	ccccagaaaa	19560
aataataat	~ catotttctt	ccaggtttta	aatatcaaac	: aaaagggatc	tgtgggccca	19620
~t-==========	a taactettaa	tagatttgat	. tttcctgcat	ttcctttatt	ttgatccagt	19680
~++==++c	a tatadaatta	r tototttaac	aggattctct	: taaaattcct	tetteagitt	19740
acctaccaa	c ttttcttat	· ccaggtttca	gtatqaactc	: cactegatia	acayayetet	19800
ctactactc	a cttataaaat	: aaatteteta	r aacatttctc	gaagtgttgt	galagigal	19860
224244724	c actactacto	: traatttqtq	, tocttactac	; atgttggcti	tialacycal	19920
taattaaaa	t taaggagtto	r tagaaaacat	: ccatgaaaaa	a acagattaa	laadacaacc	19980
atacatata	t ttaaaactaa	r aaggtactat	gagaaqqata	atcttcatac	cagaccata	20040
ctycatyta	a tttcatttat	- cantttanac	aaccacttc	ccttcccttc	accetacete	20100
cryacerga	a tasatttat	tcaccactct	ggcaccacat	cctcatccca	a gcaggatttg	20160
cgagtgcct	g tyacttigte	cottttaaa	ttctataaa	tgagaaaata	ctaggggaat	20220
ggaaggctg	t ttatttaas	a tracacoct	tagtcagtai	tatgacagag	cttttcctag	20280
gattttaaa	a statesses	t ttacayycci	ttttadddt	t tcagagacc	ctaaagccta	20340
aaaaatgtg	a tagttagtt	r totosttot	r tttaddtac	cttctaaaa	ccttttgaga	20400
tccatagac	t andergate	g telegaligi	a aatgtaatt	a gcggaaaga	a catttcaaag	
tgttaggaa	ic cacaacaga	g carecegae	_ ~~~	9-999		

						20520
actgttgttc	tgcttagact	ttctagtttg	tcttctgcca	ggcttgccgg		20580
+++aa+aacc	tratactcaa :	aagaattgac	atttaaatta	gtetetetet	Leccuty	20560
h	catecttete	tctacattct	atetetatet	ctgttagctt	accecec	20700
casatasaca	ggatatagtg	actattattt	cttcccctta	tccttcaacg	acctactet	20760
~~~~~~	taccttttt	ttttgagat	ggagtttcac	tettgilgee	caggergggr	20820
atastaataa	aatctcagct	cactocaacc	tttgcctccc	gggttcaagc	cattttttt	20880
antogacete	ccgagtaget	gggattacag	acatgcacca	ccacgccigg	Claatitigt	20940
attttcagta	gagat ggggt.	ttcaccatgt	tggtcaggct	ggtcttgaac	Lectigatette	21000
aggtgatctg	cctacctcaa	cctcccaaaq	tgcagggatt	acaggcgtga	gecactgige	21060
cotacctact	atttgccttt	ttaatctcat	gaaatgttct	cttttcttgg	Cigaagigic	21120
acttttcttg	ttgaacagca	tgcgtggtga	gtagaatgtt	ataaaaaggg	acgyacticg	21120
anattemene	garccaggtt	cctattcaac	attgcagaaa	tgetgttetg	caataggetg	21240
tatataaata	ggcaaattac	ttatctctca	gagccttatt	ggtaaggtgt	gagigalage	21300
taatttcaaa	caccttacag	aggetatete	ctaatcctgg	tagogtacct	ggctcacaga	21360
taacatttaa	aantnottot	gatgacagtc	atagctcacc	attagcalag	cyclygatec	21420
ataacaaaa	agegetgeac	atocaotato	tcttggacta	cacagggccc	LCalgaalla	21480
agaagtgotg	tttcatdadd	atagggatga	ggaaattaga	ettgetgeee	Cicacigcoi	21540
tacactacta	tectecaaqt	taatgggaac	tatgactctg	Ctttggcttg	actyccaty	21600
aagattotoa	cacagccaaa	tttattqcta	tcttagttaa	attatgctag	aacacaaaac	21660
atgaagttat	tatcaaaqta	atataatctc	agctgtaact	gagatagica	gaaactgtct	21720
atastatast	greetatetg	aaaggtagct	gagaataaac	aagaaataaa	gagaattag	21720
taggaaatat	taataacaca	aagcttttat	attttgacta	gttaagctag	licitaaaty	21840
tttccactaa	aatattcaag	tttaagggca	tagcccaggg	cagettatta	Lyaacatyat	21900
~tatttaga	aatcttacac	tttctcttaa	aagttcttgg	gaggggcarg	cyayyccaca	21960
atataaccat	aaaaccattt	gttttaaaat	aaaacccatt	tttaaaattc	coccaacaa	22020
aaaaattatt	gcaggaaaaa	atgctaaacc	tggtttttaa	etttgtaege	tatangent	22020
tccaagatgt	gctgtagcct	ggtaaccata	cagaaccata	cagaattagt	atasttasa	22140
tattgtctgc	ttacttttgc	atttggtaca	ggtataacag	ggtcgattat	atygututta	22200
agacatgact	agaaagaaat	atgtttatca	gttattattt	CEECCALCIA	tasttastaa	22260
ggctagggag	agggcttcaa	caggaattta	tatactttag	agaaaagtga	gggatcatac	22320
cccaatagta	tagatatete	aacccaataa	cacaggttgt	gcccgccccc	acaaaacttt	22380
actgtagggg	agaatctttg	caagcaacat	tetaettata	gggagccata	tgaatgataa	22440
catatgtata	ataattataa	gtcttaagtc	atcaagaaaa	agttaacttg	attaatttt	22500
tccctgatta	aaaagagaga	tgtataataa	tggataagag	actiticity	attototaat	22560
agtattaaaa	tggctaaatc	ttctttggga	tattetgact	ageaeggege	cctctttagc	22620
agatttccca	tagctgagag	ctaatcatct	cgcaacctgc	tttctacaga	ctttttttt	22680
taaaacttta	ttgtaattcc	tctaaatcct	caycullat	cattttcato	gcctgagaac	22740
tttttaacat	ttccttcctc	tgactcactc	cettegeece	tttcttttca	agtatogttg	22800
atgggtgatg	atagaattat	tetttteaca	contracatt	tcatttttat	attttcttc	22860
agctcatgtg	tgtattaact	agagaagtet	ttgaaatgga	antttctgcg	gttcttaaga	22920
tcatcaggag	g atagtttgta	gecattlact	cecaaaceca	ttaaaccato	tectetattt	22980
cctgtatcat	ttgtctcctg	aatttcactt	tatttatta	attagcaagg	gccactctct	23040
cccatcttc	t gcacccactt a tttttcttt	therapaget	. egctegceed	ttctaggaag	ttttttattq	23100
gtgttggaa	t caattcatac	ctyaaayccc	ttattttac	aaccctttgt	taataacata	23160
ctactgtta	c tatagttatt	e acceptate	gatcatttta	cttggttaca	taaggagcac	23220
tttatttaa	c ccagcatcat	. agcagcccgc	gaccaccett	totttoatto	ctgtcctaac	23280
atatatcta	a gtoctaaaaa	. ctcattagta	gegagatgaac	tateettace	tttgctgaa	23340
ctagtaccg	a cacactgaat	e etteatage	aattcataqt	tocatttoo	cattttttgt	23400
catatatat	c ccctttctca	ttaaaaaaaa	tttattttct	agactttagg	atttagagaa	23460
ttataattt	g ttccatacac	atactacta	togattatt	aggtatttt	tgactgtatt	23520
geteattt	a aataaaaag	ctttcaaga	atocaaaaa	a aaaaagctca	a aaaaacagaa	23580
ttatettig	<pre>a datadaday\ + ++++===+=1</pre>	- ctcagatag	tttaaagaa	ttttaaacat	cctaatcata	23640
~+><+++	a addccattca	a tagtacaac	tgtgaagag	c ctcatgtace	, cyctaactyy	23700
atactatat	c tacaattaa	- togattgtt	g ctgacatct	t ggccarcay	g Cayaatgege	23,00
+	+ acactacat	r ctgaaagaa:	o oottaaaat	g getgecally	g cacgggcgcc	23020
-att+act	a dotaactto	r ttccatgct	t ttctctcta	t atatgtagt	l lalaaallil	23000
	+ +++aranac	a ofctcactt	t attoctcad	g ctgagtgca	g tygtytgaac	23340
agaggt car	t acadeetta	a cototogod	c tcaaqtqaa	c ctcctgcct	e iguellulaa	24000
	va contantor	c caccatcat	o cccooctaa	a lilliciali	t titglagaga	24000
tagagate	c actatatta	c ccaggctgg	t cttggactc	a agcaatctg	c ctgtctcago	24120
-55555						

ctaccaaaat gctggattat aggtgtgaac tgccataccc aaccctataa aaatgttata 24180 ttttaaaatt taacaatata cttcatgtga atgtatggtt tttaaaatgg gtttaatagt 24240 ttattctcag ttgaagtaat tttgtttggc atttttagtg gtgtgtattt atatacgtct 24300 gattatccat atgcggtttt ccttcagcat ctgtggggat tggttttaga accaccacag 24360 ataccaaaat ctaaggtgtt caagaccctc atatagaatg ggatagtatt tgcatataac 24420 ctgtgcacta ctttaaatca tctctagatt acttataata tctaatacat tataaatgcc 24480 atgtaaatgg ttgttatact ttattttta tttgtattat tttsattgtt atattatttt 24540 taatttttat ttgttcacat atttttgatc tgtgatttgt tgaacctgca gatgtggaac 24600 tcatggatgt gaagggccag ctgcagtaaa atgaaagagc aaaaatgcaa atgtacaaag 24660 ttcaaacaaa taggaaattt aaaggcatag aatttgatag gcaattacat taaactgttg 24720 cttactttct ccagttctgg aggctagaca tccaagatca aggtgttgac agggttagtt 24840 teteccaagg cetetetece aggettgeag acageatect tetteetgtg teeteaggtg 24900 gttttttcc ctgtgcccaa gcacccctgg cactgcttcc tcttcttaga aggactagtt 24960 acactggatg actaatcctt ctacagagac tgctaaggtc ccactctgag gccctttttt 25020 aaccttaatt accacctcta agtccctctc tctgaataca gtcacagtgg gaactattag 25080 ggctttagta gactgatttg ggggaacaca cttctgtccg taacagtgcc acataaatat 25140 ctttagcagg attgattttt taaaatccct aaagatcgtg agtattgaca tgttaaggac 25200 gctttttagt gactctgtaa taagtgggtg gaagaattgg gagttaaatc catctgatgg 25260 ccaacaaaac taaacttcat gtgttttcca atacagtgtc acatgcagtt tttttgaatt 25380 atgttgagac aaggcaattt tcagctaaat gttctttaga agctaatgtt tgaagatatt 25440 aaatatagat taaattotga aatgtagtti toattotgta ottittigoaa gagaagttigo 25500 ctttttgatg actctggcca attgttattt taaaagtaaa tgctctttct cccgatttga 25560 ttgtggcagc atggaggaat ctatgtaaag cgcagtgcca aatttaacga gaaagagatg 25620 cgaaacaagt tgcagagcta cgtggacgca ggaactccag taagagccta cccgttttta 25680 tttttcttac cagctctcag tttctaaatt taagaattaa attaaaatct aagaattgtt 25740 ttgacaatgt attttcccat gtgtaattac taattcaggg ttatgctgag gtaacagaaa 25800 controlled acagging agging agging geometric against acagain against acagging agging agging geometric against acagging agging agging geometric against agging agging agging geometric against acagging agging gtcctttgct ggtcagtgga ccttaaagag gaataaaaag agcatttggt gtcgttcaga gtctataaat agaactaact gcattttaac ctgacattta agctagttta caagctcatc 25980 ttacttcttg tcttctttag tatcagattt ggttttagaa gcagcaactg ttttctgtta 26040 gtgcaaattt tgaatgtctt acatgtacag aaaaaccaaa aaaggatgaa tctctacaaa 26100 tgttaaatca ttcagtgtaa ataatatttt ataaaacttt attccacaaa agtggggaga 26160 gttcaatctg ctttgtatag aatgctgatt gctgccaaag gcttttcccc tggttccctc 26220 cggagacaaa gcaccatgat caccggggcg acttgggctt tetetttcag tacatgacat 26280 gtgctcagaa gcttagctcg tgtgcacagg ctttcccttt cctttctggc tccctccctc 26340 tgtcttccct cctctcctt tgccctcccc tcaccagggg tcctgggcag cagctggagc 26400 tcatggtgaa ggaagaattc ttcatggtca gctggcgaag tgcctggtgt gagcattgtt 26460 tattcacatg cotottctag gtgtttttac attagaacat tgcatctgtt ttgggcatgt 26520 gttgggtgac agaagcagaa tggaatgaga tgaacagtga ccctttatcc tgttatagct 26580 aaccettgag aaccaagett ggtgtettea aagggtetgt ttagtetgaa acagtgtggt 26640 gaatttgggc agaattgtgg tcattgcatg taggtctcca aaagacagaa taagttggta 26700 atatggttta tcgacttttt acaaaaaaaa tttaaaaaatc atgaatttat accttaaaat 26760 gtccatccca cttctctcc agctgtccag tcaccccagc aatggatgac tgctgtggag 26820 ttccttctgt gtcctgctgt gggcattgta tatatgaagc aaatgaagat agctgccttt 26880 tgggtgatgt tggcatccta tgcacagtgg tcccttgctt ttttgccccc atgaatatag 26940 ctgccagtgg cgctagggct gaaaaaatca gctctttaca cttgtcatgt gtcttgttta 27000 tgtggctgcc ttcgtgagtt tcttcttgtt tttggtttgc agcagtttaa gtatcatata 27060 tctgagtgtc atttaaaaat ttttacctgg attggtcctc tgagcttgga tctatgattt 27120 ggtgtctgtt attaattttg gaaatttctt tgctcttatt tccttaaata ttattcctac 27180 cccagtcttt cttctccagt tatgtttgtg ttggttcatt tctcgctgtt ctttagttct tagatgcatt attegtttt tgttggtttt tttttaaatt tttttttta cgccccctcc 27300 cttttttctt tttgtgttac attttggata atttctgttg acccaccttt gagttcatgg 27360 attetteett tggetgtgtt gagtetactg gtgagecagt ttaaggeact etteatetet 27420 gctactgcgt gtttcattcc tcacatttcc ctttgaccct gtttcatagt ttccatctct 27480 gtgctagtgt atctatctga tcataaagct tagtcacgtt ttccagttga acctttatca 27540 ttttattata cttgcagttc tcttaaattc cctgcttgat aattccaaca tctgggccat 27600 atctgagtct gcaaattttg attactttat ctcttcagat tgtgctttat cttgcctttg 27660 tcatacttcc taagattttg cctaacgctg ggcctttttt gtaagacagg agaaatggag gcaagttgtc ttgatacctg gaaatggata gacttgtctt tctgcttggc ttttagtgtt 27780

gaggagtgga	gtcagtccac	tgaggaggtg	cactgcattt	gggttttgct	catgtgcttt	27840
ttctcacagc	ttcaggtttc	tgtagaactc	attactttgt	ttgtaggttg	gggatgtcct	27900
cccgctagag	cttttcctca	gtgtctattt	cacactcagc	gttttcacat	agcaccttgg	27960
agtggctctc	ttctttatgc	ctttccccac	tatacttctt	ggatacttgt	tactgaactc	28020
tcgctagttt	ggtggtagaa	ggagagggaa	gggaagtgtc	ttttcattct	tagggagaat	28080
ctcaggggtg	gagccttctc	tgatcctgcc	ttgcttctgg	ctgtaagtct	gtgcccagta	28140
tgtattcctg	cctttactaa	gagtttttcc	ctgttctctt	cacccagcct	catcgagtat	28200
tcatccgtgc	cccatgggta	gcagggtttt	gttgcccctg	ttcatcagtt	tcaggctgct	28260
gttccatagg	aaaggtagaa	agaaggatgt	gggctgggcc	ctgagccctt	cccacagggc	28320
tacttttccc	teccacaage	ctacatccag	tcttccctga	ccgcagtgtg	ttttctttt	28380
tetttatett	gtgagtacac	aggaggtctg	tgggtcgagc	ctgtgaaatg	tgctgcattc	28440
teettatate	tgtagcccag	gggttcgtct	gttccactgg	ctcatacttg	gctttctgca	28500
aaattgataa	aatttttagc	taaattcttt	ttactggtat	ctgttacatt	ggcccccaac	28560
taaacaacca	cttgcatctt	gtttctcctt	tgagttttcc	atctttcctt	agacttttgg	28620
gttagttggt	tgccttgcaa	ccttgcagct	ctctgaaggg	totaagaaaa	gtcatgaatc	28680
tacagettgt	cagtgttgtt	gttgttgtag	ggttggcagt	agtattcctt	cagcattcta	28740
catacttaat	ggaagccgcc	tcccattttt	ggttaataaa	tttcaaaact	tggaacaatg	28800
ttagatttac	aaaaacgtca	gaaagaacag	agtgttcctg	tttattcttt	atatagettt	28860
tttttttt	ttttttt	agttggagtc	teggtetgte	acccaggetg	gagtgcagtg	28920
gcacgatett	ggctcactgc	aacctctgcc	tcacqqqttc	aagcaatctc	ctgcctcage	28980
ctcctgagta	gctgggatta	caggcgtgca	ccaccatacc	caactaattt	ttgtattttt	29040
agtagagaga	gaatttcacc	atgttggcca	aactaatctc	caastcctga	cctcttgatc	29100
cacccacctc	adcccccgc	agtgctggga	ttataggtgt	gagccaccac	gcccagcctt	29160
cttcatctag	ctttaacatc	taatgttgac	atcttacata	acatootata	tatttgtcaa	29220
aactaagaaa	taaacattgg	taccacacta	ttaattgtac	tacagatttt	tattcagact	29280
ttaccagatt	ttccactaat	gtcctttttc	tottctaaaa	tacaatccag	aatagataca	29340
aatccattca	actroagtgt	tttaaattat	totttttcat	tatatgaagt	actatataat	29400
ttttgtcasa	totattattt	tggttttaat	cttcaagctt	arctttattt	ctttaagtga	29460
taaagggata	atttaaaagg	tgtgttgggt	tatttcagtg	cctaaagtct	tatctaaatc	29520
acttattttc	tactattett	gcttatggta	ctttctttcc	ttatttactt	tottatcttc	29580
ctttactact	agetatattt	ggttaagtta	tttgtggaaa	tcagttgaag	cctcaggtgg	29640
gagtgtcttt	ctccaaaaa	catttctacc	tattttaact	gggccctta	aggeteetet	29700
aggggccccc	ccecggagaa	gagattctga	attaaaaata	aactgagcca	ttcaggcagt	29760
agegraggee	ttacagatac	acgtgagacc	tactcacctc	tcatttactt	tcaccctgag	29820
gcagccaggg	ttaatattta	gttcacttgt	ctgattctct	cttcacagtt	ctattagaag	29880
agragageet	tttaatttat	gtgcccttca	tettateact	cttotaaato	aaagttctgt	29940
tttataatta	cttggcctct	actgtgtttg	cttaatttca	gtcttaacat	cttgccaact	30000
attacgeeca	tttaaaataa	tgttatatcc	acctttttaa	attattttca	gtaggaaggt	30060
trattrasat	aacctactct	ggttatgggc	taccaccata	acctactat	tttttataaa	30120
casasttoca	accettetata	ttcctagcgc	agtgtggata	acagactage	aggttcaaga	30180
gassatasta	aggagettte	actgtaaggt	cactatacca	actcaaattt	ctaagaatct	30240
ggccgcgccg	ttcatttctt	aatatgtacg	ccctgtgaga	ggcogggcot	cttactcaga	30300
ttattataat	tetttett	ctgaaggtga	attaagtaag	tgacatggta	gaatatgtta	30360
agtcaacttt	catataactt	actagttctc	atgaatgtat	tccatgattg	tatcagttct	30420
tattcactat	tagtatttaa	gaaatgcaga	attttatttc	aaaaaatata	tttgtattat	30480
aagttataaa	cagcacccaa	tccataatta	ttactagae	aatacagtat	tttcttaagg	30540
aagttgtgaa	ttataatac	aaatgaagca	tatttqtqat	aaaaataact	aatagaagtc	30600
attttattag	actatoacct	agtaaaactt	atorcacaaa	categagact	taacactttt	30660
tettecaget	ttcacttaac	ttccttttca	gataggagg	agectagtag	ataagagtat	30720
taattttaaa	attacattca	ggtttaaatc	ccadatette	totttaatct	ttattttatt	30780
teagatagat	tttctcctt	acttgctata	acttatacat	cagtacttgc	cacttcaatt	30840
ttatggtagat	ceeeggaca	ttettteett	aaacctcacg	aaccaacctc	tactaacttc	30900
taacyttaty	gayayacyyc	ctttacctct	ctcagccttc	acaccaattaa	agggagttag	30960
gacettacta	tarattara	tttgctttaa	addagecte	taactaattt	gatgttttat	31020
ggeettgete	cygactayga	ccatatcage	pagagagata	ttttaattta	taatcattca	31020
tatattanet	. caaacttcct	ttttaattct	ctttaaggetg	ttttccttc	catccactca	31140
tractet	. yaaytaycac	acctagettt	tracctacct	taactttaa	catacettee	31200
tanatanasa	. aytyyaaayy	attgatgtaa	antrararar	atocaactot	teettteact	31260
ccactaagcc	acciding Ct	tagggttatt	aatteees	atttcaatat	tattatatat	31320
ggaacgctta	, grayerateg	aaaaaaaat =a	nenesensers	aracarrara	acaggecate	31320
attacacac	, gyaaacccaa	caccacattt	atcaattaaa	tttataatat	tatatgggtg	31440
arrygaycag	Leagaacaca	Lacyacacti	uccuactada	. cccgccaccc	cacacgggcg	27440

32700					ttttctaca	
32040	atgattatgt	атстаадтас	tttaaaatct	catggattta	attcttgtgc	tacgatacat
34980	aagtgcttgt	аассаадссд	atatctgaga	attattacat	ctttcatgct	ccttaattca
34920	agtttcttct	taggaaatga	tattatctct	ttagtgtaca	attccttgaa	gaatgtattc
3₹860	ttgaaggaat	tgttttacag	aagatactt	cttcctggaa	tttagtttt	taataacatt
34800	ttttcaaaga	даасааасаа	εεασαεαεεε	аадсааааас	ccsscgtggt	catttgctgc
34740	dcredrcedd	agreerrea	, ಪರ್ವವಿತಿನವಾಡಿ	tataatccag	ಶಡಿದ್ದಿಕಾರತಡಡಿ	ttttccaga
34680	racccacas	cccscsdscd	במכבכבשבמב '	ತ ತ್ತಾರದ ಪ್ರಕರ್ಧ	cacacaaaat	ತತ್ತಿಂತಿಕಾಗಿ
34620	асаасаасаа	сссеядасся	ddrrradcrr	agcattagta	atactctgtt	астдтаддаа
94560		асавасасас	ссясягдсяя	ರಿತಆಂದಿತಕರಿ	ταοοοοετ	gcatttgtt
34500	כבבשבבמבב	rargrecara	ragercacge	atttaaaata	ರ್ವತರಿಂತರು	ತಂದತಿತಿಕರಿ
34440	rddcrdgccr	адаггсгса	ששבמבבבב	atataattag	aattattagc	ttaattagtg
34380				cattgagget		
34320				дасссава		rarraarscr rarraarscr
34260	בשכבבבבשם	ccaaatgagg	atgattttca	ממלללמכלמל	scstddtddd	agaacgtttc
34200		derrrassr	ссгдгдгдээ	aaataggaat	dsdccdcdcd	гдсяссяяяд
34140				ttatataaca		בפבבבבבב
34080	בשבככבכבב	rccagcrda	rearrassed	aattgattac	аасттадаст	гвадсдгава
34020	ccraagcaga	בבבבפכבב	васагссед	дсасавсссс	αθτττταστ	tcaagctagg
33960	aactcagttc	בשמבבבבה		дгсаатсааа		creartreed
33900		вазавассе				atgageteet
33840					rscrarctas	drrdrcrdad
33780		decdrecede			tcatccagat	
33720		дсесерадс			tatctaaaga	attataccag
09955	rrragaracc			ratttgcgtg	ತನಿರ್ವಧರಣ	двасдвасва
33600		адерсрава			_	atacatataa
33240			בשששששבששב	בשבבשב	ttaataataa	aaattattgt
33480	ttatacatat			дагсгадгая		ttttctacat
33420	rrareccesc	гессоста		ataataagga		trectrorad
33360	כבבככבמבבב			αρααραρρ		atttcatga
33300	сасававсас		стдаатагаа			гдагадссся
33240		raaatgactg		адаватавтс	таставсстд	ccacatgagt
33180		асссавдася		_		atgeetteet
33120	gaargaarrc			сататаадаа	acataataag	agttcctgga
33060	ροσερεσος	гдгаггагг		ccatgaaggg		tattcttt
33000	сгагггаага		ತತ್ತು ಕ್ರಾ ಡ್ಡ್ ಕ್ರಾಡ್ಡ್ ಕ್ರಾಡ್ಡ್		_	כבבבבכבבב
32940	αςαςςταττ		totoatoatt	ctaaattttt	attgcccttc	tcctagcccc
32880		tettttatee	aagcactcat	ctttagggaa	tattcctcat	atcctttcca
32820				acctcctagt	rrcrddcrød	ttgctaaatt
32760				crrarccrr		ссссадавас
32700				atattoctta		catctagtaa
37640	מכננמנננמנ	асстедства	саататаста	rttgacttt	taggcatata	tttgttgacc
32580	csagttttg	aaaaaaacct	gatgacattt	tactattgta	cagtttttga	ctaagaattt
32520	адагагагсс	שבכבבבבפכ	agatttgcac	tttaatgtac	rdrrrdrad	tatctcagtg
32460	כננכנננשננ	rracttaggr	cracarttgt	tgaacgtagc	rrrddarrca	ddarrdaarc
32400					בבבבמשבבמב	
32340					בבבכבמבמבמ	
32280	getetetag	вассассса	rrcrrrcaa	aatttgttct	ttgtcttcta .	aatagtgtag
32220					atgccaatac	
35760					gttatcccta	
32100	дассдаасас	дваасааса	cccaatgttt	attatttccc	ratagaaaag	taaccatcac
32040	ссаасасадд	בכפבפבבבבב	ಇರಿತದಿಂದರಿತ	atatggtgta	raatgtttc :	בפבבפבפב
31980					בבבבבבבם	
31920					rrrargrarcc .	
37860					ttttgaagaa	
31800	tttatacatt	בבבבבבשבב	ಶರ್ವಿತಂಡ್ರಿಕ್ಕಡ	שבכבבבבבם	tgtgaatagt i	ttttcctatt
37740	grtacttacc	ttttgcatct	tgtacaagtt	agaatgttcc	greactecta	ಶನಿಶ್ವರತಿ
31680	tgataaaatg	ಶತದಿಂದಿತತದಿಂತ	aagctcagta	aatttccgtg .	aaaaaaatgc ,	cttcaatggg
37620	tgtcataaac	derdcedddr	agacttgctc	tggcaccaat	arragasasa	agcacatact
37260	ಶಂರ್ದಿಡಿತಡಿದ್ದರಿ	rdacacadad	гассдавата	attgtgagat	aatatgaaat a	cagatataat
37200	atcacaataa	ನಿತನಿರ್ವ ಿ ರಿ	agtaacatca	caattacaat	сассссааа <i>•</i>	caattcatgg

38760	מבכבמבשבבב	rdffccddg	ttaatgtgga	פבבבבבבב	רבשבבבבבשכ	ttccctgtta
38700	ddeedeedeg	regreecerg	כככככבבבב	cattggtttc	trttgtatta	ttccctttgt
38640	raaacarcrc	cctcttcctt	aatatcacct	grtgaagata	aattttctt	aatggagtat
38280	ತಿರ್ವತಿಕಿಕಿ ರ	actttattgt	aataatttta	ataaccagga	ttataccttt	ttgacattat
38220	cacataactt	tgttaacttt	ttgtcatatt	tttgtgtcag	catcgtatca	acctccaaag
38460	rraaccactg	cacttctcac	aaatgaatcc	cactctccag		eeceededee
38400	decerrace	адаадсседд	'೧೯೩೯ ದಿ೧೮ ಆಶರ	aagtctatag	tgaaataact	
38340	LEGCABAEGE	receccade	್ದಿ ದೇವಿದ್ದರೆ ಬಿಡ್ಡು	ccttacttct,	cagtcatctg	atcgttgcct
38280	cccrarccc	arrecgrera			atggcaacgt	
38220	дававассс			ಶರ್ಡಿದೇರಿ	aaactgctga	tcttdgaag
38160	дасседдасе	ddagccccc			atatocottt	agtgcaccca
38100	crcrddaggd	адстатаст			ttcacattca	ttttgttctc
38040	rrrcacggar		ccffffggcf		agcagttctt	
37980	dccdsdccss	tttcatttca	вдссвддсся	tctgcagcca	crascraccr	aagatcaatg
37920		ttagcagtgg		כבמבבבמבבב		ttcctcagcc
09848	ρασσερασσερ	αρροφέρος	αρασραασορ	gcccatccca	cccddddfcf	craccceaac
37800		αααοορρο	dcsdcttdcs	τααοοοετος	tetettteet	atttctaatc
07778	ερεραασεα		tcactcgctc			catatagctg
08975	tegeceacae		catctggaaa		gcttttacca	ccagtgctct
37620	במבמכבמממב	rrcdsddrcc	craracraa	drcrdddarc	ccraractra	aa rrcaadar
09575		crarecadaa			rrraraasaa	dcracsarcc
37500	carraaceaa	ಡಿತತ್ಕಾರತಿ ಡಿತತ್ಕಾರಿಕ	ttactcagct	ಡಿತತ್ಕಡಡಿಡಿಕ್ಕ	rdagaagtct	ctcttacatt
37440	cagatttett	taaaacacca	crrddrddcr	ಡಿಡಿದ್ದಿ ಆಕರ್	ctctcaggca	ccrdrddcrd
08575	מבמבבבמבב	rraagaaggr	actetegete	rddfaacggc	cagctttgtg	agctagtgct
37320	сваасаадаа	dssagcatag	caacttctga	actttgaagc	ttaaatgaat	gaagtagaaa
37260	сасаассааа	במבבכבבכב	aacaaatgtt	cacacacttg	acaatgacc	ttcaagagat
37200	дватавдава	ತತತತಡಿಡಿದರ್ಶ	ದಿತ್ತಾತ್ತಿಕ್ಕಡಿಡಿಡಿ	acataaacta	attagaaaag	gracrrcara
37140	таттасавая	atatacttaa	саддасасса	ttttataata	gctaataata	caaatggcaa
37080	atttagttgg	atatgaacaa	ctcaaacatc	atttttatgt	aataaaaatt	gacttaactg
37020	rrrcrrdra	aagcaatttt .	ataagctaca	ಶಶсಶಶತದಿದಿತ	ತತತ್ವದನಿಗುತ್ತ	gagttctttc
0969€	agagatttga	raacasrrar :	tgaataacac	attcataggg	tttccatgtt .	atacctagaa
00698	ದಿದ್ದಿತಿತಿಂದ	тдгстатааг ;	tttaatgatg	ಕ್ಷಕ್ಷದ್ವರ್ಥಕ್ಷ	ctcttaaaag	
36840	atacttaagc	cdssddfdsc s	ragttgctag		aattgccatg g	
08498	cratagragr	вадсавассс о	ggacccatta	rctgccacct	tgtatcttaa	tggaattaaa
36720	araagggarg	tctaatggtt :	tatttggctt	ttgtttagg		ttaactaatt
39990	αααεεεαεε				adcactatag '	cagaagtgca
36600	radaarccra		ದಿಡಿಕಡಿಡಿಕಡಿಕ		cttataggaa a	
0₹59€	ctgatagcag	aaagaacaaa 1	ttattcacaa	tgcctaagtt .		
36480	rdfcffddfa	tcatatttct i	atgcatacac .		aagatagaat 1	
36420	grcrcaaact	ccaaaaggat g	cffgactggc (gagacttog (ttgtccatta :	atttatacaa
36360	crcrddaccc	стстдсства с	ctccatgtct i	gerreereee ,	actcgcaaaa g	actqctgatg
00898	מבבבככבבם	ctttagccat o	aaatccatgg o	gecttectet :	teteatggtg g	acttttatt
36240	gattaatttc	רבבממבמבבב מ	agtgtcaggt '	בקרבת ברככ :	cracracara ,	teatttetee
08T9E	cettetget	cttccagcc t	tectageete o	gacagttact '	ccaactctct g	cattatcotc
36120	αδοράρρορο	scrccarda g	s parocotodo	atttqaaatt	בבבבקקבכבק :	ccrcaattta
09098	creactetge	tttccagta c	tgtaagcact i	tacctttgt ;	aatctgtagt o	accactotod
36000	свавасава	ttaaacagaa c	teacattta	: בדבֿשַכפכדב	ככפבבקבכבכ ו	rrraceata
32840	csgttgctc	tectgtcatt t	agacatgacg i	ctgtctaat	cattgettee t	בבבבספסבבב
32880	gccagctctc	tececeatet g	tecagttggt (ಚರ್ಚನಿತ್ವದ್ದು ಚ	agtgetteag a	traadtaadd :
35820	scatgiccta	tetttaaac s	atgecagttt g	נדנננפפפפ	rareaceate c	tascatcatt
09725	sataggtata	gcagatatgt c	g epacaaga g	ttacaaggt g	ಕಡೆಡಿಡಿಡಿಡರ್ಡ್ ಚ	s elesesette
32700	tritagrita	בדבבמבבממ ב	aattttaa i	acatgaaac i	ardrdacrad r	Seeptiens
32640	_	cctaccaat c	מבמבבממכ ו	receptosas,	receteday;	heestootes
32280		taatggaac t	resatactat c	ומממככבשבם פ	a cattacago a	s proprietto
32270		reacteact o	atattaata	toatctccs t	. totostitien	gettgaetgt g
32₹60	gretttgaa	tettagagg t	refactata i	researda c	aactatacad s	ccaaattggg a
32400	gcaatttaa	saataaqcta t	rradracac o	s elecolises	s spettttede s	cttaatatcc a
32340	מַמַבְפַבְנַמַבְרָ	crrcaacaa t	cedteato i	reacetedas c	2	tracacaaag i
32280	rrradadada	caadttet a	t pistasaapp	Tienedien	2 Becoesces 2 Entritrit	gragatotaa a
32220		r seerssess	s ettototeo:	ימממטרטטי ומדדממדדמד ו	o spincings.	tgacagacct a
32760	gessettaa	retataton t	> e44e444406	= 100106669;	- 1nenn11nee	- 400000000

45450

45360

45300

07227

42180

45150

45060

45000

 $0 \pm 6 \pm 5$

088TF

41820

09LT#

00LT#

079T7

41280

072T7

09717

00PTP

41340

08ZIF

41550

09TT#

OOTIF

OFOLF

0860₹

0260Þ

0980₹

40800

07407

0890₹

40620

0950ħ

0050¥

07707

40380

40320

40560

40500

07T07

0800₺

400S0

09668

39900

39840

39780

39720

09968

39600

39540

39480

39420

9868

39300

39240

39180

33750

39060

τροσορίασες ασοτρέσερα στροσορίας τολοκολοσή διερορίας εροροφορος οσόρροτρο τροσερορες ορεροσορες ορεσερερορ

110

tatotttgca tggcaggtgg gtgcacgcag gcatgctctg catgtottcc tottggccat cacgtgccac cgcgcccagc tecgtgttet ettgttee etgecteetg etetecaet occesectoce caacttocag gotaaagcag tootcotgag tagttaggae tacagacata etetetetet titetitetg etitetitet attetetete etecetetet teetetetet

tececeteta gitatggigi ggettatet aegegitetg gageagaage etagteacaa

caccacagag tgatattttg aaactttgga cttcataaag ttggatgagc tccagtagca tttctgtctt tatttcctta ctaacttact tggatgccag taattagttg ttttagtgaa agctatttt ttäaaacatt catgataatt cattcottt tatgitttaa aaatactagc

гдадаттттт дегдсаадте гдедаатсаа тетаастдсс сседссседд ддастаааде वस्तुवस्तुवस्तु प्रतिवस्तिक प्रतिस्तिवत्तुवर वस्त्वस्तुर प्रत्यवत्तुर्धे पुर्पपुर्पुद्वित्वुद्धे

cagatacgtg cttgtgggaa tctttgtctt tcccacca ccctgcattt taaaacctct

agaggattga aatogtatgt teatacaaaa gettgtteae etgeageett catatgeaat tggataattg cagtagtoco ccagottaag aacottttaa aaatatgtca gatatacoca аддавдавда давдавсвая ссстевссс садтаддавс стальдавту ддагтгдтгс קאקכנומני בנוכנישנות מדנקנכת הבינונות וכתה במושה שמשמשקה שמש שמשמשקה שב gagaggccag tagaaaatct agactagttt tttatagtct attttcatgt cacttattga tgtgtgggac agtcccacca tgtaatagct gttcttcctt actcagctac tttccctcca

tattcagcca taaaaaggaa tgaagcactg agtcctgcag ccacacagat gaacctcaga дгосатдаад сааттаатад дтааасаааа тдтдатотдт тоасасадтд даатастаас tectatgaat gtteatagea geattattea taatageeaa agtatggatg caaceeaat

acticcaaca cygigactiy tacatytaya tattiyatca atalalayca aatyaatati ддсяствата гредссерд ссудутста стадагрува усуберссус гадуссява ಶಿರ್ವಿಚಿತ್ರದಲ್ಲಿ ಇರ್ಲಿಂಡಿಂಡಿಗೆ ಪುರ್ವಿಸಿ ಪ್ರಾಂಡಿಸಿಕ್ಕಾರ ಪ್ರಾಂಡಿಸಿಕ್ಕಾರ ಪ್ರವಿಚಿತ್ರ ಪ್ರಾಂಡಿಸಿಕ್ಕಾರ ಪ್ರವಿಚಿತ್ರ ಪ್ರವಿಸಿ ಪ್ರವಿಚಿತ್ರ ಪ್ರವಿಚಿತ್ರ ಪ್ರವಿಚಿತ್ರ ಪ್ರಕ್ಷ ಪ್ರವಿಚಿತ್ರ ಪ್ರವಿಚಿತ್ರ ಪ್ರವಿಚಿತ್ರ ಪ್ರವಿಚಿತ್ರ ಪ್ರತಿಸಿಕ್ಕ ಪ್ರವಿಚಿತ್ರ ಪ್ರವಿಸಿಕ್ಕ ಪ್ರವಿಚಿತ್ರ ಪ್ರವಿಸಿಕ್ಕ ಪ್ರವಿಚಿತ್ರ ಪ್ರವಿಚಿತ್ರ ಪ್ರವಿಚಿತ್ರ ಪ್ರವಿಸಿಕ್ಕ ಪ್ರವಿಸಿಕ್ಗ tecatgetga gegaaagaag ecagaaacag gaggecatgt getgtgtgae tgtattteta

titgatigca tgaagaatta titagatgca atttatgatg ttacggtggt ttatgaaggg ctaggccttg cagtattaaa acatgtgcta acaccacgaa taaaggcaac tcacgttgct greasacage atattgtgga ggtgttttgt aactaattaa tgacggcact gaattgactt ggttgtggaa ttaatagatt ctgttgtgtg tgtttgaggg aatttaaaaa taatttagat पुत्रस्थववटव पुत्रवेववपुष्टव वर्षपुवयुष्टिय प्रस्टिचववट स्वयुवयुप्टर वर्ष्यवस्य

tattaagatg taacagtgga gatttcatta gtcctgcaaa gtgtggtatt tcttggctgt aagtaggtta agtgtacttt tttcctccat tacatttacc cggtatattt ttcaagatgt cctgaaatgc ctgtacacgg tatatacagt gcacatgttt atgtagaatt cagttttaca वयवविवरतेवरते वयतेवितवतेत वयतेववेत्तव ट्येवटवर्ष ट्येवेरवयते व्यवहर्ष व्यवदेष व्यवहर्ष

agaattttaa gtgagcagga taacttagtt aaactaacca aacatagtgt tagctgttag occasacatg ttaggtttta agaacctttg agctattgtc agatataacc agaaaaaat attaaattig ggigicatig igacaagaag aaaigcagit aagigigaco iiitiiitic gtactgagtg cacgttatta gtcagtgctg cccttttgct gtatttttg taaaatattt асасттавая дувавустст утававудуя удавудству выдавус всусстууса cgtgtgagte etgtggacte accaattate attaatecag cetetteta eteaagtte

tectgagtag etggaactae atgegtgtge caccatgeat gactaatatt tgtattttta gcaatcttgg ctcactgcaa cctccgcctc ccacgttcac acgattctcc tgccttagcc tttcdttttt tdttttdaga caagagtttc gctctgtcgc ccaggctgga gtgcagtggc asastatotg tttgagotgg ggttgagaga gaacactagg cttcatgggg ttttttgtt अवस्वस्वर्य स्ट्यात्वस्य स्वव्यत्स्य स्वव्यात्रस्य विस्तर्या विस्तर्याहरू

tagacattat taaagcatte agaagtgage aaggatagaa attattetge ceaacettae קפרטניטרני שקפפבברנים במקפקנפבלם, לונדנד בפקפקנפב בבבבבבפק aatatgagce tecageaegt titacatgga aacceteace tgaageatte gtetgaagtt dagccactat gtatgtcagt gtgcttgtat cagtaggatc tactgagggc agatgttcaa ccacgcacct cggcctccca aatgagcttt gtgttttac ctcatcagct gtttggggtt dradatatgg gatttcacct tgttggccag gctggtctca aactccttac crcaggtgat

पुर्टायुर्वतत्वत रेत्रेत्वत्वत्व वर्ष्यवपुष्टित टाट्यपुर्टार वर्ष्यवुवेत्वत्व टार्टार्पुवेटर catctttgga tgttagggat ttttccaaag ttcagtgaga ttatagttgt caaatgatta acgtaggeet teragaegta gtaetgtgea eegttaeatt atetaaeaet gtetgtgtgt

0809₹	tagaattgca	ccraccraca	aataacagcc	ತರತತಡಿತಿತ್ರಕ	tetteatett	agacttagtt
46020	arcarcra	cradaccrra	catggctacg	atagccccac		ತ ೦೦೭
09657	radarceras	rarcccracc	ctcctctcat	rrarcacara	gcttccact	acattgaact
00657	arddcarrrd	аасстсстаа	rcagtgttag	tttcctccct	ataccactta	adsagtattt
07857	гааасстага	адтасттат	tggatatata	tttgcaaatg	taagatagtt	atgetteaae
08457	эвсэгдээсг	ttactcatgt	agggegaete	aaggtgctat	tagattccct	tatttttcc
45720	гасставав	ававсавсав	,ಕರ್ಮರಇರ್ತನ್ರ	actagaaatt	rrradraccr	atcagtatac
09957	acceaagege	בבבכבכבמבב	:	tdfacgaata	acctgtgacc	сಶದಿಂತತದಿಗತನ
00957	σεεααεεεα		rdccrrcada		ddaatgtddd	tcatcttgtt
00557	crecedder	cagcagrear	сдедаесева	tdffcattga	ddctatatag	ttttttaaa
08757	rrccdraacr	grararacac	dracatgrea	ataaacagct	crrarararr	ttcatatttt
02757	ρεσεσερες	rararcrers	ασερερεα	ttaaagatga	tapgtattt	tatgtgcctt
09857	rrddadadad	гагсгэгага	בבבבבפפ	ttaaatcctc	atatttcct	aacaatttgg
42300	atttcctt	caatttggat	deddfaacee	cacctgtcca	tcbacaagtt	ಇದ್ದಿಡಿತಿಡಿತಿಕ್ಕಲ
0023V 07257	ccsaagaaga	tacagataac		dddgccgddd	taaaaaaaa	agaatccgcc
45180	rdacagagea		ccactggact	raagategtg	£dcsd£dsdc	ಶಡಿದಲಾಡಿದಲ್ಲ
45120	rassccada		ರ್ನಿದಿತಡಿದ್ದಾರಿ	ractcaggag	taattccagc	raggcacctg
09057		aaattagccg		catctctact	ddrdsssrcc	tgaccaacac
00057	ragaccageg			ರ್ವದಿರಿದರಿಗಿದ್ದ	rrdddardcr	tcccagcatt
0003V 07677	acgegtgtaa		ದಿ ಶಡಿದ್ದದಡಿದ್ದ		מבבבבבפב	ttctattaat
08877	tactttccat	acaacatatt		tctaccagta	ttaagtttt	atgatcctag
44820		attetatttg		ರ್ಥಕ್ಷತಿಗಳಿಂದ	taaattgtgt	tgtaatgtaa
09477		בשבבבבבבם	בבבבבפשב	actattgatt	tagtttgatt	cdaaacagtg
00277	ರಿಂದಿರುತ್ತದೆ	agargreare	taatcatgga		tcatggttca	gaacattctc
07977		дгатааатса		_		arccrrccca
08577	agtgtccctt			daggtettat		attcttgtag
44520	rdcracagga	· _		ದಿಡಿತತಿದ್ದಿದ್ದಕ		cdcrcrdcsd
09777	dcraccadda		_		tgattgttga .	rradcracad
09000	rrarcagas			_	catggcccag	gttgatatgc
00777	ttggtagatt					tttaatattc
44280	ragtaggeta			_	drddccscsr.	taagtacttt
44220	tggaatactg			aagttgtgtt	catctgatat	aattatatga
02277	agreacted	ревттаретт		εααεςεααεε		atatgaattg
00177	tcaattgtt			_	בבמבבבבבב	tetgetttea
05077	ttgaaaaac				acacgtttt (tataataata
08687	gcacataaga					ragatttagc
43920	attcacattt			creddaaacd .		caaacctttg
02827	caatgtagta		שבשכבבבמכב	aggcatttt :	aagtattatc :	darasastrr
43800.	aggereageg			adarradcaa s		בבבבבבבב
00857		дааастаса ;	aagtgcaaat i	ಕ್ಷದಿ ಕ್ಷಕ್ಷಕ್ಷ ಕ್ಷಕ್ಷಕ್ಷ	aacctcattc a	tttcacattt ;
08927	acaatagaca	_	ಕ್ಷತಿತಿತಿಕ್ಕಾಗಿದ್ದ ಕ್ಷ	staactagaa a	вадагдасгд :	aattgataag :
43620		rfdcdraaag		ಚಿತ್ತತ್ವು ಚಿತ್ರಗಳ	срадарсара б	tatttgcaac
0.9554		cdaaagacaa s	ಕ್ಷಕ್ಷದ ಕ್ಷಕ್ಷಕ್ಷದ ಕ್ಷ	saccagagta s	aggaaaaata o	tctgaattaa :
00567	าาอาธิอออาร	cedicessan c	מבבשבכשכב ו	gaaataaaa g	ataccataaa g	acttacccag a
43440	Baccccccd	בבמשששששבב	ברתבכבבב	cattacaaa 1	ctatgtgtat t	tagtcactgc (
43380	Adcreradaa	adcrascrar a	gctacatgtg :	בבשבככבכב	cactgccctg c	tecaacetee
43320	າຄິຄິຄິລວວຄົລລ	ardedirace i	asascerect :	:daddaagcc :	ataagtgcct t	cadecgeate a
43260	בכבכבבבככ	ας εξασίες τ	cradcacadd a	lracradese ,	gtgttaca g	cactgatggc g
43200		ctcttgcagt a		:draccrrcr	agcatagcac t	s opotitions
05165		creceadad r		cctcatcca s	בכבבבבבם כ	
43080		saateetget t				gcaccccatg o
43020		ceddcedsce c				recreerara e
09627	מככר בכמכר:	שבכבכבת ב	: ೧೮೫೩ ಕ		ctdccatcgg s	ccrcrarrac c
42900		gagereatgg g		iagaacccct t	screerarer s	
05825		gretaaatg		agactecee a	daaaaacaga t	tataaccaat
08724	าวถ้วถียยวยเ	craagatta s	taaacatag i	:מבבממבבבש	caagattgt c	agacagttca t
42720	ובבבכשששכ	בשבבבבבם פ	gttacttac c	tatcgaagg a	stetttaaa c	בממבמבבבבב כ
42660	deddogege	cracagasa s	:כככבככבב כ	attetacea t	atatottet t	caaatttaag t
00927	ובבמבשבמ	:בבמבבמכ כ	ttttcttaa t	cdataacca <u>c</u>	tecettecta g	geetteeett e
42540	figeracerg	cerceace e	:מבכבבככבב כ	acagtgaaa c	tracgaaatg c	aattaaaact c
42480	_	tgtaatgca t	gtttagaat a	agatgtaaa t	בבבבכמבבב ב	aactaaggtg t

'N2DOCID: <MO ____8835844455">

0ħL6₱	בכבמככמממב	deccradic	agcatcott	cedddecrrd	attcatatat	atgccattt
08967	ассавасас	catatgetat	ಡಿಡಿಇಡಿಡಿಳ್ಳಾದ	asaggatttg	555555555	tttcaaaaa
49620	gagaacacc	dccrddragc	дссвасьссв	cratcattgt	ರ್ಧರ್ವಚಿತ್ರ	ತತಡಿರ್ದಧಿತತ ಿ
09567	ccdagcccdd	cdøddørcøc	ಡಿಡಿಂಂಡಿ ಇಡಿಡಿಗೆ	ctactcagga	dragreceag	gcdfdcacct
00567	aarrageeag	аваатасааа	drcrcracad	gcaaaccct	ಡಿಡಿದೇಶಕದರ್ಶದ	aagacagcct
0030V 07767	ccgdggdccc	rederedade	ddceddceeg	dddaddcfaa	ccagcacttt	gcatgtaatc
08864	daraderegr	adceadceae	,ಕಾತನಾತನವರು	asatactcag	cacagaccaa	ttccaccaag
49320	carccargga	arggarrcca	greestaree,	gtagtgagca,	csdcccsddr	drrrrdaga
49260	gaataaaggc	sacccacaa	כבמככבבכב	cttctttac	ταστατττο	gagttaacta
49200	ತ್ವರಿಭಿತತಿತಿಕ	creatgraar	ctttacaaat	gctatcataa	darsrarrcr	tattatatcg
07167	במבבבבשמשב	בבבכבמבבבמ	ασασερεσοσ	ತರ್ಇತಿಕಿತಿಕಿ	tgatttataa	crødddrard
08067	acgarrareg	ттттавася	ccccddaca	аттадаата	ttttataagt	greeraerta
02067	במבשבבבששב	ававдастая	drrcrcadaa	atgaaaatat	atgtattgac	taagatctac
09687		ccatggaaga		ccattcagga	arrcagcgag	gatgtaatgt
00687		attgactctt			taacaatgaa	tgtaataggt
07887	cdarraaaag	ceddcdaaga	traactttac	ccccccddd	ttttaattaa	saracsarar
08784	cracresc	בסבבסבבבס	cratcattgt	gctcaagaaa	tgtactttct	
48720	שבבמבבבשב	атттевдося	atattcagaa	gragreatt	agtgtcttct	
09987	deedgegeded	ಡಿಳಾಡಡಿಡಿಡಿದ್ದ	grrggaara	tgttatattt	gacatgatgc	caracsadara
00987	accardodga	adaareaaa	ಶಶದಿಂತತಡಿಗಿತ	aracaceaes	rrdddccadr	ααεςεααςεα
00507	racrararas	ತತತ್ವದರ್ವರ್ಥ	caggtcacat	acctacctgt	agatgcccgg	ttccagtgtc
08787	садададасс	васвавсасв	taaaaacaac	caaattaaat	gataacatag	aatactacgt
02787	rrrcradas	attgatagta	atctacttgc	csgtgggccc	gaacttgtta	cagtctaaaa
09887	regracasar	taaactttac	gategtgttt	atgcatgtct	atcaaatttt	ataaattcaa
48300	ששבשכשבבבב	redecessa	аасддссаса	aaccaatgta	aacaatttga	cagtatagtc
48240	aactagattt	screedarre	cacttccgga	acagttccag	draadraaca	tcaaagataa
08187	carredagg	derdeardaa	£dagaaga£g	caagaacata	೧೭೭೮ರಿತಕಿ	aaaagatgt
48120	carecases	tcacattgat	asattcatat	gaatgtccaa	tetetgeaaa	tgttagaatt
09087	בבבהבבבב	aaagaattat	tagtaaaaaa	agttctaaaa	tatgaccatg	ttttacttta
0008₽	ננפננננפנ	recrarrace	שכבבבבבב	tcatttaca	tacttttct	aaggaatctg
0 7 6 <i>L</i> 7	בשבבבבכב	ששכבבבבפב	סבבבבבבב	grttataatc	ccaaccacca	೦೭೬ರಿತಿತಿತಿ
08874	tctacttcag	aacatgctct	tagaaattat	gragttgact	ttatgtaatt	acttaatgac
47820	tattgattta	catttggtac	tgaaaaataa	agttgtcaca	tagcactaaa	gggctgtcta
09 <i>LL</i> 7	tectattect	ddccccsdcc	acttagagga	ardadtatgg .	crrcagagag	rracadarcc
00 <i>LL</i> 7	caaceac	ಡಿಡಿಡ≎	cactaggatg	acaggtgaca	csccstddtg.	aggcagcagt
01941	ρος αλαρα				agccagacca i	tgccatcgtg
085 <i>L</i> 7	ಡಿರ್ದಿಂತಡಿಡಿತಡಿಡಿ	crdrdrdccs			gaggattatc	
₹ 1250	arcadacasc	daagctgtgt i	rcagtgacga i	ratatagosa	בבתבבתבתם.	tececeaagt
09 7 L7		ragagagag i			arāacssras ,	
00717	aagtatcagg	aaaatatggc :	ataatatgtt i	s fattacobet	בבממבבכבבב ו	agtaataatt
0 ⊅ €∠₱	gctattccac	adddcreatd i	agttgtttac .	s padbaaapag	ತಿತಿದ್ದ ಕಿತ್ತಾರ್ದ ಕ	ttttcaaat
₹7280	actgtctcct	gctgaaaag a	tgcaatcgag g	catacettg '	aataatagcc	tcttaaggtg
47220	ctccdttctt	agtttctagt (೧೮೪೪ ನಿರಿದಿ ಅಭಿಕ್ರಾಣ ನಿರ್ವಹಿಸಿ	gageceeeta o	rarccacagg ;	cattettace
09TLÐ	aatatgttaa	aaggaactta i	. בבבקברני	gdaatatte ;	ttaaatctta g	tagtcagctc
00TLÞ	ctccctacct	cteggeatac	sagcetecea o	ceaaaqqtct o	caatttgagg c	tititeseses
07017	cacgtctttg	segestojji	gaattcatca	rtadagitata g	recreased c	agacaagtot
0869₹	ದಿತತದಿಡತದೀತತ	בַּבַבּמבמממבר מּ	aacgggaagt g	s Esapeddaa	tacadadat :	deagagcagt
0769₹	ctgaagtcaa	gagcicgaca (stgttactaa	agatactac s	tttaatcacc 1	acagggtatt .
0989₺	grāraagaar	coctggaagt :	accadttac i	esettatta.	t ttacacata 1	מברברמת במב
0089₹	taaaatcttt	aatgaaaaat i	adacaaaaat a	addaaaatta a	atactet s	tattcagatg :
0 7 197	jttetttgat	rtttaagca g	taaaaatac t	יםרדמדמכרמ כ	ccfagggaac	1 2112120101
08997	pteceacetg	gotacctda	cadadateca	CCCCGGGGG	Treesagatt :	deddaeadtt i
₹6620	gcacagagag	geaddogoadg	tatadactt o	atccctact t	t tettettete	ttgaaatctt
09⊊9₹	SEGERADAS	caagcactc o	σεσετατατα ο	acatadtos:	. setopeteet	tasascacta (
0059₺	Satttgtggc	graagaadr s	ataadatad c	soeteeeps:	· tototetti	gaatetgtet
07797	reeddeded	aggeaggaa a	atteadtas s	idatetadat s	s portoseptic	arcryrarge a
08897	gattacttaa	rdaccaddt c	rectaegata t	eseccadat s	s spitiosopes	atctgtatgc o
46320	ertecttaca	בפרמברככשב כ	daaatacaa t	, 115550151;	. VELECTOR	cactgoctct t
76260	casacscra	raccacate c	g googeasace agradadada	2 JAJJOSEPES 2 JAJANJEN4:	אאממממרממ כ	gracgritta
00297	teregaega	saccettt c	y googeances.	ייםנפטנטמט פ יישרמטניטאר י	אממרממררמ c	gagatcaaat g
07T97	dependent:	1 TIEDEDEEF.	t noontaonee	= =====================================		

Ell

23400	draacagcac	гасадсстад	accactgcac	ccatggacat	rracearaea	gaggttgaga
07889	ffcagcccag	ನಿಡಿತದಿಡಿತ್ಕ ಂ ಡಿಂ	ಡಿಡಿ ದ್ ಡಿ ಇಡಿದ್ದಿದ್ದ	ctactcggga	trtgtcccag	cactgagtag
23280	rddrddcdra	гядссяддся	taaaaaaaat	acaaaaaat	gaacctgtct	aacatatega
23220	csgcctgggc	адгесаадас	aataaggcca	gatagaagga	actgatcata	atctaactag
09185	csatcsagtg	σααροραραρ	ttctttgaat	cataattata	tteetttgtt	getttattt
23700	ggttacataa	дсгеветаб	caacaattt	tttagagtgg	attttaaaa	tttccatgtg
23040	arargerea	gragrerece.	aracatactg,	ctcttacacc	atctggctaa	aaaagtgccg
23980	drrddrdssr	ааттаастаа	aaatgcctta ,	gaaacatgtc	tgatctcagt	aatcattaag
22920	gcaattotgt	aatttgcttg	daccarrarr	taagatgcat	ctgcaattta	ataaattcat
25860	deredderge	cagtatatta	aactgttaag	aaataatcaa	aaatacacta	agttttaaaa
22800	сгггдадсаа	tttccctaag	tcacaatatt	cctgcattct	aatgactggc	aaataggctt
22740	ccagtacata	caacagtete	taaaattcac	ttttgaccc	tgtttaacat	atggctgtgc
22680	actgtctcta	aaaatcttc	ававававар	atgaaagtaa	dededeetad	tcagaaaagg
22620	tttatgtttt	rdrdcardcr	atgccgtcga	ctttataaga	attaatttt	cagcttagtt
22560	tactgattta	tttgtattat	taaacttttg	acatatcatt	aagtettte	acctacctgt
22200	agagccgtat	сстававста	tgaattattt	gatactgact	aacctgtatt	gettetect
25440	aaaattagaa	авасдеседд	ttttgtatca	ttgaaaaaa	aatacagtcg	aggtctataa
22380		tatactaatc	ttcatagaac	aatgttaggt	agatcacttg	gacctgcata
22320	ваддевдссе	tcattttaac	ggattctaac	daattatgt	caatactgtt	cr c t c d g d d d
22260	сааасссдаа	dddrardrrc	tgaatataat	taatatatt	tacagggatt	cfffdffaag
22200	aarrccargr	аасссадасс	сдсавадсав	ttctcttacc	aattgtaatt	cctcatgctg
22140	argaaaarrr	ggtaaaactt	гсддддсягд	ggcctaataa	ggatatagtt	tggacactca
22080	cacacatcca	ತಡಿತಾರದಿರುವ	tgtgtacatt	t daat t ggt t	rcararaara	cfffgaagag
22020	caagcaccar	taaactcttt	rdccadadaa	deffecees	grarracact	ttaagtgact
09615	reedddeese	aaataatgat	аааасдааад	catgataagc	ccagttggat	attaaacata
00615	agradeceae	дааасасасс	ttttgctgca	ರ್ಥಿತಿಂದರಿಗಳು	rcracttact	agttgacatt
07875	ragecerea	gatttcctca	tttaaacttc	сдедерава		agtgtgaact
08712		cccscdddad		ttagccatgg	atcacgattt	сಶದಿದೆಶತ್ತು
27720		ttagagcaag	atgtategee	acacattaat	radradador	ccattaggtt
09915		tgaacgaatt	tcgtcatgac	cagacttaat	cagtgaaata	cttctcaagc
00915		arcrararcs :		rdcsdsddcs	caccagctac	ಡಿಳಡಿಗಡಿಡಿದ್ದರ್ಧ
07515		rdaactett		dracracce		aagaaatgca
08775		rrdcacgtag	dcrcedcaac	csataatact	radatcccat	drectreecr.
27420		radarreces	rareararaa	מבכמכמבבבכ	tcattcagaa i	cacagaatat
09815		ρασορασερο	αραρασορες.	atttactage	rrcraccrad :	ccfdcadaca
27300		даттттавд :	aagttcacaa	arradacccc	grgaaccttg :	aagagtcggt :
27240	севесевес	בבבבלבבבב	בבכבתבבבת .	сдревадар	tgcttgaaaa i	araccereed .
08115	ככבבבתבכב	מבבשתבשכבב מ	aagctacgct i	rtgaagattt	teceagettt	actttaaatt
27770	ccatagcaga	crcrdattca :	aaaattgtga 🤞	aacgaagttt .	aaaatatgag s	aaaatggtat a
09015	gracacadg	rdcacaccct :	rrdcrraarr .	rttattgaaa .	atttctaata i	tgcaaaataa s
00019	שרברבשבשבב	gattttacto :	affattagot g	artgtattgc	acatgatttt :	rcrdrrradc :
07609	addsssssdc	adrdrrcard :	aagggtaaaa	grrcaaagga 🤞	gtgtcaggat g	tatttaggaa g
08805	gageggeed	caagtcagtg a	כבבשבבבכבש כ	catctttg	agttaatatt i	tettgeaget a
20870	gegereagea	carracrada g	ccaagctct :	<i>ತರ್ಮಕಡಿತರತ್ತರ</i>	atttattaca a	gttatttca
09405	משכבבבבכב	sagedtattt c	מבבששבבבם כ	sattttgag :	cagatggctg (ttttgcaagt (
00702	gggcgcgcg	ಚಿತ್ರದ ಕ್ಷಣ್ಣ	acaacttgag a	: daactaatg	catttccta :	cctgctctgt o
07909	:ddsdrrrcr	явавустава 1	್ವದಿದಿಶಿಇಡಿದಿದ್ದ 🤉	ргсададсга и	caactttagc :	cctddttgta (
08505	gcaacata	ಕರ್ಕಡಿರ್ದಿಂದರ ನ	getgatatge :	:drdacdaaa	caatagaatt i	gaagattgga t
20250	בבמבבמשבב	eaataaagc c	rraaggagra ,	ccdaattta :	saaattgttt o	cacatgacat
09705	radcadcra	בבמבכבשבבב ב	ardrdcraca i	ಚಿತ್ರದ ಕ್ಷಣ್ಣಿಗಳು	gacrarcrc :	atagacaagt a
00705	carraaage	erdedddeeg c	σεςεααεςας ο	radagaccer :	acctggatat a	gtatgtgaac s
07505	haaddaadcr	sccdsrdcrd 6	sarderrard s	:dactgcagg (taagtggtt t	aatgttgatc t
20280	בבבשככשבכ	rcsagaaga c	caaattaagt s	rgttaatte o	ctdddssss 6	aaaaagattt c
20220	ಚಿತ್ರಕ್ಷಕ್ಷ	caccagaic c	atttatgag i	:gcttataga :	gccatagga t	tttccattct g
09105	זבכבששכבב	adrdacraaa s	gaaatactga s	:ttttgaatt :	gattgatat t	aatagctacc t
00105	ccedcdrdr	cataacttt t	csaggascer <u>e</u>	: מכרב שברבר מ	ccgtattca t	tetggatgtt t
07005	saaccaagtt	sacttctatt a	gaagactaga s	igatattta g	agcatattaa a	ctaatcatag s
08667	מכשבלבשבש	וכשבשבשפש ב	בבמששבשבב ב	saccacctg t	saatgeetta s	catcctcata a
02667	rdecesesed	raagiiiaaa s	sactttatca s	tcactgtaa s	:dragarced :	attagocatt t
0986₺	acedcadar c	ಗಡಿಡಿತಡಿತ್ತದ್ದ ನ	:afaaaaagg s	iddadcrcca i	:tddcddggr:	tettgaett c
00867	ceddecera	ctaagaacc g	1ಡಿಡಿ ಆದ ದಿನ್ನಡ ಕ	cdatacaga g	sagccccctg t	atcctgggac c

BM2DOCID: <MO___8832644A2_I_>

5₹0 780 750 90	tgtgtttca ggtgattcac	agttaagttc agttaatctgg	ತ್ವರತ್ವರತ್ವ ನಿರ್ವಹಿಸ್ತರತ್ವು	ataaagaggg tttacaactt	gactgttcgc gggtttggag tgcacaaaag cacataaaca	grcaggagct tggcaaacat
			,			<210> 180 <211> 2000 <211> DNA <213> Homo
26520	craracrada	drdrraddod	τροσταστος	caaatttgaa	actcacttca	sdsdcrdrdc
09799	ಶಶರಡಿದಶಡಿದಡಿ	ddrdrrcraa	drdrrcccsa	gattaaggta	crasaarraa	ttaaagattt
00799	gctatcttcc	gaagttaata	aaatgagaaa	atctactaga	gaatttggtc	dragrodrra
07895	atatagaatt	ataatttca	ttactgtgaa	rdatttggtt	rraccrraar	tattgtttcc
26280	gttatgatgt	ccatacgtat	cagtaaaata	atgaatacat	cctggaatga	ggagatgtgt
26220	ರ್ಥಿತರಿತಕರ್	gcaaatgtaa	aaactgttta	בבנכבמבמבכ	aactaaagaa	agagaactag
09199	tatcagttga	ссааастааа	aagaacactt	tggattctta	ccaatgtatt	cacacctade
00199	cdcdagccac	ggattacagg	caaagtgctg	crrddccrcc	atccacccgc	acctcagatg
07099			catgttggcc	dadarrede	tagtagagac	בבבמבפבבב
08655	tggctaattt	ccaccacacc			crcccaagta	
02655					ctcaatctcg	
09855		tgctctgttg			111111111	tatttagatt
00855		cttagatcag			aatggaaaca	ttgtttatat
07455		taactactgt			tttatcattg	atccccatgt
08955	•	aatteceace				コココロコココココ
22620		gcacccatct	ttecteaact		catgaccacc	agttacttc
09555		tttcddacaa	tottgatagt	_		tgctaaacat
00555	·	Energered			attttttt	
07755		ctaaaatgtg tgattgtgag	tatttgctgt		ttgctaaacc	
08899		tttaacacc				
22330		ttcagggaaa	cttcaaagag			_
2250		crectarra	100118888	22222222	232333333	
22500 227 4 0		cradacascs				deddrrdced
08055	ccfdcsddcd					acctgtaatc
22030		cceddcdrar		actaaaaata	· · ·	cdfddfdsss
09675		tttgagacca		ggarcactrg		rrrdddaddc
00675		cacacctara				tgcataactt
07875		αερααεαερα		aaatcccaga ,	rrarcaaggr	aacagttacc
08742	agtttgtacc	adrrrradgr :	catactgcca		aaaagtaatg	
07175		aaccccatgc ,			cadacstcat	
09979	gcagtgcact	tagtaaacaa	rttgtagtga		greatattt	
00975	atacggetat	gtattgocaa a			gataaagttt	
07575		tggtgataaa			gcaataaagg	
08779	attggtgcct	ttttgagatt i	- gacttgtgt	gaataaagtc	ggtatttca	atetettat
2 4 450	aactctttt	atacttgact :	ರ್ಷಕ್ರಿಗೆ ಕ್ರಾಪ್ತಿಕ್ಕ್ ಕ್ರಿಪ್	tatateteea :	attgaaattc .	gggaaacac
24360		atgttcatgt (aatttataa	rararrerae :	atatgtaaga '	ttatatotoa
00879	grecatgatt	attattattq (r recepteedd '	adatcadttt i	ttgatgaaac	piepiieiii
24240	caacaaaatd	aacaaactct (raatatotoo	asadassars	ctaatatact (1126126521
08179	aatadaaadc	aaccacttca a	ttttaaass :	saccessors and the terminal te	crasacsada a	ayecatatate
24120		statotatac i	eleileilo.	, respectedes	ctttccaat gaggtcaaag	ratetestat
09075		rangossass Hatteetata	, presetpese	s antaththi	gctaatttac ;	Tenentite
00075		. grancarcar	. senetinetr	caatttaas.	ttcatgtaat 1	ררארכשששרש
23840		. Dieiteines	. 244226664	1 1155010561	ttagaatcaa :	cargeaarea
08865		y pespeddaya.	י בכפטינטמט	ישרמשמבנשם ו	gaattacgt :	caaagaactt
23820 23160		. Specie de dite	יירכארבשכם י	1 168006661	בבְכֵּכבבכבב	aagaccccc
00752		. Seeqbeeqbe		geraceco i	aaagctacaa	rarcraaraa
07985		בבבכבבממכ י	gaggggggg	מבכספים	tagtgtttag :	בבמבבבבפמ
08252					מרדרקרקרק ו	
23250			яваваесдей з	secoracias	aactcccaga	ರ್ಥ ಚಿತ್ರಕ್ಕೆ
09789		בפרפפפפררנ פ	я се	್ದರಿಕ್ಕಾರಿಗಳ	ctcttagaaa a	ದಿ ತರಿತ್ತಾಂದಿ
00003	- , , ,	4 4 4 4 5 5 5 5 5 5		•	. , ,	

09ST	חששב בשששחש	ggdgdggccg	atatcagttg	тссааассаа	. ಇತ್ತುತ್ತು ಇತ್ತು ಇತ್ತು	ttggattctt
	cccaatgtat	662226262	dcdcdsdccs		ccaaagtgct	ccffddccfc
1200		gacctcagat	tegaaateet		ccatgttggc	
7440	gatccacccg	•		tecaceacae	ತರತ್ತರವಿಧ್ಯಕ್ಷ	
7380	ttagtagaga	ttttgtatti			aacctccatt	
7350	cctcccaagt	cctgcctcag				
1560	gctcaatctc	gagtacagtg		ברשכבכרשבר	ಶನಿಶ್ವದನ್ನು ಕ್ರಮ್ಮ	
1500	בסבבבבבב	gtatttagat	್ವ ಚಿತ್ರವಾಗ್ಯ ಚಿತ್ರಕ್ಷ	gcttagatca	ತನಕಿರತಿನಂದ ಕಲ್ಪ	
OPTI	raatggaaac	tttgtttata	tgagagatca	graactactg	gaaaatagca	ggctagagtg
1080	rrrarcarr	carccccard	ccsccdctdc	gaattcccac	ತರಿತತತತ್ತಾರದ	aagatggtgg
1020	rattgcagag	reffesage	ссстагаада	tgcacccatc	tttcctcaac	ctttattctt
	ccatgaccac	בבבכשבבבשכ	attaaccagt		recedarad	cscsddrcsd
096		gagttactt	gagacatgca		cararacr	tattcaagct
006	crarcradar	ttgctaaaca	ggaggarrg			адгагггаа
048	tettetggt	_	graataagt		reatteegetg	caattcagtt
08 <i>L</i>	tatittetga	_			ברכברשבשבש	тааааатаса
720	cttgctaaac		caatgtgaga	2626611116	•	atatgtattt
099		credrrrdag			acttcaaaga	• • • • • •
009		ttacctgtaa		tetectgitt	aaaaaattcc	999999999
075	999999999	ctccgtatta	ಶಶದಿರದಿತದಿತ	ccfdddcdac	racscrccag	atggcgtcat
087		ವಿಡಿಇವಿರ್ವಧದಿ <u>c</u> ಆ			ಶನಿನೀತನಿತ್ತ	taggaggatg
420		cacctgtaat	೯ ವಿರ್ವಿ ವಿರ್ವತಿ	ರ್ಡಿಚಿತಿದ್ದರಿಗೆ	acaaaaatta	tactaaaaat
098		ಇಂಡ್ರಿಂಡಿಕಿಕ	ಶರ್ಡಿಂಡಿರು	drrrdadacc	dødd£cøddø	tggatcactt
	ccdsddcssd		_		ರ್ಡಿಡಿಡಿಡಿಡಿಡ	ffccfdccdd
300				ααττααταοτ		taaatcccag
240	ttaaataaat	•	2276227622	2661111662	dcatactgcc	descedess
180	cttatcaagg	caacagttac	ccacagegta	622222222	2022222	teteatetea
TSO	taaagtaat			Dienare	5-57-57275	
09	gedddegreg	taagcetett ·	accadtacac	atagtaaaca	ағғғағяаға	
					•	T8T <005>
					sabiass	omoH <£12>
						<212> DNA
						<211> 7601
						<210> 181
0007					tgagaagatg	ರ್ಡಿಂಡಿಂಡಿಕಡಿಂ
2000	าวควาวคิตก็ก็	ಡಿಳಡಿಂಡಿಂಳಡಿಡಿಂ	ನಿನೇವನಿನಿದ್ದಾರೆ ನಿನೇವನಿನಿದ್ದಾರೆ	ಶಡಿದ್ದಾಡಡಿಡಿ	dccdcdscrd	ccacccacca
0861		cadacccada c		åãã⊊ããcããã		ç ವಿವಿವಿದಿದ್ದ ವಿತ
1920	oeecessoos adatantaa	3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3				caacacacac
1860				2		gagcatccct
1800		מכבבת כבבת				rdccdcddgd
ዐቓ፟፝ዾፒ	satggggtgt			cddceddddd		
0897	adcacarrar					
1620	aatcaagaga	ವಿತಿತಿದ್ದರಿಗಳು				
0951	adccdddrr 19	desdradeed 6	addagcatgc g	ವಿರ್ವಾಚಕ್ರಗಳ ಚ		
0057	ccaccacct	ccarcaarc c	ttcacccagt t			
0777	ಚಿತ್ರಕ್ಷಣೆ	19ರಡಿದೆ ಆಡಿದ ಆಡಿ	carccaaccc 6	accacacaca (ರ್ದರವಿ ಕಿರ್ದಿಗೆ	
1380	ccsddcsdc	addracacar r	rarasacsac s	cccsddddcr ;		
7350	ρεσεραστο	מכככשכבבב כ	araacgiraa c	radacctccg :	cracsagasr	
	acaacctgg	იენიენონე	acasccada c	daderdeece e	ವಿವಿವಿಶ್ವರ್ಧ <u>ಇ</u>	araeceaer
1500	ישמשטטטטענע ישמארי	מרכיב ב ו במיט ו	ენენნენნღნ	ემომმილმომ გ	cactcagagc ;	adcracraca
7200	6700126006	s onengagos:	200000000	רמשרורות מש	всясддяесе	ರ್ವದಿದಿದಿಡ್ಡರಾಣ
0 7 TT	laacsaasaa) Jenenenane	e esacanasas estrutasasese	. 6224344654 . 622434654	adaerarc	าดวาววายาย
1080	ettactotte:	Teptettee.	t Dieneseter	, 2262626242 116306242	ggaatgaag	เติดกากการ
TOSO	agtatggaa	rtctaataaa c	peeeeptot-	t sentnenser	neenteeen Deenteeen	רר במשר בבב י
096	caagtccaa	s aspectedes s	tattaaaaa s	scasaactcc t	#C9899999	
006	cdcaactac	icacatgtgc a	ttatggatt s	aadtatoctt t	aaggetaeaa a	continuate s
0₽8	cccttcgta	icagttatcc a	satcaataaa s	cetgetttaa c	greergraga 1	adatccaaaa (
087	ಡಿಡಿತಡಿಡಿತತ	109ಡಿ0099ಡಿ ಡಿ	sattttacc a	satattaaac c	aaaaggtta o	actatccttg a
720	ragagraca	adcarcer r	raaagcarr c	seratatate:	t atatatata i	tatatatata 1
099	ರ್ವಿಚಿತ್ರಗಳ	1.ದಿದಿದಿದಿತಡಿದಿದೆ ಕ	erddrddrd a	aggtttccat t	aaggtagcac a	s stooptootp
	רחשששרשכני	משכררכככש כ	accedares a	accarccrg g	ಕಿರತಿಕಿತಿಗಳ ಕ	ccatgtataa :
009	226466624	1 127761761	ירושההשמות מ	ישממרכשבכש ב	בבבקבכככפ	rresecrade ,
075	settnisni	4 464454454		1 224524226	graatgert	בכבבבשכבם
087	egatotoe.	e masessesses o soffsetto:	t 12515557.	- 2222422421 - 22224212121	sagtgettet t	n paggarant
420	adadaadct.		n spantaspor	o opogogoes.	+ 10112010e-	משחחשששחחשר מ
360	soo šoo stat.	:ככמדממכנד ד	cececa c		addadcatt a	. Tennesense
300	tctcagtta	agatgtctg q	tetagtata g	cacscscta t	בבבבמבכבבב נ	, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,

911

	•	

0912	ccdsddcddc	ಡಿದೇ ರ್ಧದೆ ಆರ್ಡಿ			•	ಡಿಡಿ ಂಡಿಡಿಂ್ಡರ್ಡಿ
2700			raffcccagg	ತಡಿರ್ಡಿದಿತ ರ್ಧಿಡಿ		ccttctggaa
7640		crddrrcrcg	7 7 7	cccdccdcdd	creaeggeet	cdfdccdcag
2580	ccddcdgdcg	cactcgcagc		dcdcffdcss	ಡಿಡಿಂಂಡಿಡಿಡಿದಿದ್ದ	opooootooo
2520	£ dcd£dcdad		ಡಿಳಡಿದಡಿದಡಿದ	ಡಿಡಿರ್ದೆ ಆಕರಿಗ್ಗ	aactacaccg	crrccrcdag
2460	τααταστοτι			ತರದಿತ್ತರದಿರ್ಧ	csgcgcgtgg	ccacccatsc
2400	rdardcccgc	crcrccacac		cctggacgct _,	tacctgctgg	ggcgccacc
2340	ρασεααασες	agegteetgt		ctatgcgcta	cacacgtact	ccrddrdcrc
2280	ρασεασεαες	ರ್ದಿದ್ದಿಕ್ಕಾಗಿದ್ದ	cddcfddcds	ಡಿಡಿ ಂಡಿಡಿತರ್ವಿಂಡಿ	dddcsscdcs	aradecaeae
2220	rdøddrccrd	cddcddfddc	ಡಿಡಿ cಡಿಡಿ cಡಿಡಿc	cacagtgatt		caacaacaa
ST 00	ದಿದಿದ್ದದಿದ್ದಿದ್ದ	ರಿದ್ದರಿದ್ದರಿದ್ದರ	ccatgctggg	ccccdcddc	daarccaaa	acasacacc carcacacac
S 700	ಶಡಿದ್ದರ್ಥದಿ	caacasacsa	ctttccaagg	døddedfdec	ccddcsadcc	cttcaccccg
2040	dcsccccdad	ಶದೀತತಡಿಗಳ	darcacacaa	dacaagcacg	acgcccataa	aacacgtatt
1980	dadsaccrra	cttattaccc	csddsddcdc	dreddeese	rrcdcsdrcd	ccagccagca
1920	agccaatcag	csagggaccg	draecasadac	agcatgcgtc	acgeeageae	cagctctcgc
098T	ggdctcatcc	acttcttccc	ccaccacctc	ctcacgcgtg	acttagtcgg	gracreesed
7800	tgtctccacg	agcatcctcg		cactacaget	cacgegaga	dacccrccad
07LT	agetectect		grecataace			tcattctgca
089T	dadcacagat	atgcacgtgg	tgtgcttcct	cgtgaaaggt	grassagree	
1620	dascccsacc		ನೆತ್ತಾರ್ಣವಾಗಿ ಇವಿಡಿನಂತ್ರಾತ್ತು			gaaacaagtt
72 0 0	gaaagattta	traaggett				ggaatactt
7200	acgeacagta		gaagagaact		ctattctagt	
7440	acatgtatag			agggcagtga		
7380	aagttttgc		_		_	
7350	trgtatagaa					
7560	cagaataagt ttaaaacgta	• •				
1500	tecaaggeaa		ರಿದಿದ್ದಿರುವ			
1140	cttatgaage				_	
1080	graaactect					
1020	rgccagcgaa					aacctgtgtt .
096	gcaataaag		•			
006	gagataagg	_				tettetase .
087 040	actoctttct				בכבקבכשקבב מ	ctcttatagc o
720	grassagaca	_			בבשכבבבכ מ	ccaccaacag (
099	ratctgtgac			:dcgfcffcg ;		
009	cascageece		gaactgacag '	ερααροτασερ		
075	crarctarc	_ ~		gaactcttgt 1		
087	ายวยวรยาวา	בכככבבכבבב ו	arraacresa ,	gaccttcag	gaattteeg s	
420	ccaacagaga	atgicatagg c	catttgcaga a			
360	acattgtcaa				caggtacct t	
300	: ದಿದಿದಿತ್ತಾರಿದಿತ	acttatggt t	gaaattgaa s		satttetaca g	ttacaaatgt s
240	greertree	aaagccaca t	ctdataaggg a	ggttataat o	greedaca c	ccccsdsdsc c
180	: ಆ ಕ್ಷರತ್ರಕ್ಷಣಗಳಿಗೆ	сстдатавав с	בכנככנכנכנ ו	atettaetg i	atteadaac c	tgagtacaga t
150	tgtcaatcc	cactaaggt t	geretgraa t	caccacatg		ctccaattca c
09	ಸ್ವಾತ್ತ ಕ್ಷಾತ್ತ ಕ್ಷ್ಮಾತ್ತ ಕ್ಷ್ಮಾತ್ತ ಕ್ಷ್ಮಾತ್ತ ಕ್ಷ್ಮಾತ್ತ ಕ್ಷ್ಮಾತ್ತ ಕ್ಷ್ಮಾತ್ತ ಕ್ಷ್ಮಾತ್ತ ಕ್ಷ್ಮಾತ್ತ ಕ್ಷ್ಮಾತ್ತ ಕ್ಷ್	:drsdsddrc s	atggaccacc t	ומדמכפדמככ פ	atatata c	tacagetate t
						781 <000>
					0057	<pre><5773 <5777</pre>
					2488	<222> <222> <229
					entnoei	<550> <513> Wns wn
					ຣແ[ແວຣ	<212> DNA
						<513> DNV
						<210> 182
			5 677757676	6 166 21161 6	า อาอธิกออกา	acaaatttga a
1061		. Esessenes	e vebveevoor	n papaatanth	# 5400411618	ддаггааддг а
1860	actcacttc	isassacrara c	yerareter c	אממאררממר פ אמנהרממר פ	s asegagasas.	catctactag a
7800	crasaarra	t diebeseit	מומרפוממט ו	מוממונבננ מ	reacrycada a	בבמשבבב
7.70	gessessi	o oppostanto.	ין לבבשבילבים.	ccaracgra r	cagtaaaat a	аасдаасаса с
0891	TTTTTTTTT	- 6-6-2-6-66	g ragaager g	дсявагдга в	ааастуггг а	σερεσείας ο
1620	240000400					

```
tggagtgaga ggcctgggta taattccttt ttctttgtca cactgtagca gttctgcttc
09
                                                              €8T <00⊅>
                                                            <223> exon8
                                                     <222> 25520..26016
                                                            <551> exon
                                                            <223> exon7
                                                     <222> 23387..23510
                                                             <221> exon
                                                            9uoxə < 277>
                                                     02612..68712 <222>
                                                            <221> exon
                                                            <523> exon5
                                                     <222> 19822 .. 19912
                                                             <521> exon
                                                            <223> exon4
                                                     <222> 14621..14710
                                                             <SSJ> exou
                                                            <223> exon3
                                                     16721..2791 <222>
                                                             <221> exon
                                                            <523> exon2
                                                       8282..5828 <222>
                                                             <551> exon
                                                                  <550>
                                                     <513> Wus musculus
                                                              <SIS> DNY
                                                            05678 <112>
                                                               <210> 183
                  addidadddc ciccitttgc tattcaaaag cagattgtgt ggacattgca
0557
       гадададсяд гаггсаадса дсгагасса ассааггпса сггааададу дададдуда
0057
       aggogiacgo caccactigo ctataatott actigtaatg gttttagaat atgigcacag
0777
       дссяддсгдд ссгодаястс адаястосас стдссготдс стсссаядгд сгдддагдая
0887.
       tgttttttg agacagggtt tctctgtgta gtcctggctg tcctggaact cactctgtag
4350
       toaccagota tagggataat cttttgtttg tttttgtgg tttttgtttg tttgtttt
092₹
       agtaattage tgateaetgg teacaagagt ttgagatgtg agettgtett etgeettagg
4200
       acagtitaca tatgagggag aaatgtggtt aggcagtaat atggatcaaa ataaaatcaa
0717
       atgiciggat taaactctit tagttatatg aaatttaaaa cggattcatg gcggtactga
080₽
       dificacadad atcigocigo cicigicico agagigitag gactaaaggo aigtacogoc
4050
       caggatotea cactgtagte ttggtgggea ggaactetat gtagacetea caggecteaa
0968
       давававату адгістіста тістітссва тавідадстс тавававава адвадсвавс
3900
       ctaagtgatt ttaggggettt taagttaacg cettaaaact aagattaagg agaagagag
3840
       acttgagggg agtcaagtgc aaaccttatt acccccacc caggctacag cagctgttt
3780
       3720
       accoctcaga cgctcaggga cotctactca agtgccacot atattcttgc tgcagagacg
 3660
       садаадддас аадасадсдд ддадсасаас гтааадгус гуассттуса сагуасаадг
3600
       ttactaagtt tgtcaagtga atgtatggtc taatcgtgga taagtattta atttcactag
 3240
       ggatttcatt cagaatttt aggaacccc acactgatgg tttcaaacct ccctcttact
 3480
       tagggaatag gaaaatcact gttgcacatc aaggtttctg aaaaattgac ttttagaata
 3450
       attatgtetg aggtagggaa ttaagacete tetgaateae tatetttte aatgttttee
3360
       tetgigitag gggggtactt tggcgactgg atccettact gagttttgcg ceagttggag
 3300
       ctgtttgggt tctcccagat gtatctatga ctttcccccc satttctcag ttctttcat
 3240
        acateceata aaaacetaca gttgtactte ettttggtt efeagtttea aagaagaget
 3180
        agattgaact ctgaaaacaa taaagagtag aaagctotoo taatgtgaat togttatatg
 3750
       cerderactd trictrict terrecane cerecetece ecceecee egecegigg
 3060
       αιτοιοίτος σοιτασαοία σαοιτοααία τασασοίσατο τοιταίοασα εισοιτίτε
 3000
        додаддоссь адддаддава сдаасась ддсьасаьа ддадосьсас сдсодаьаса
 0767
        cactdaccaa agtgacagct gtgaaggacg actoctgoot tgggttooto cogggtgaaa
 2880
        gatgetggga etggetteat ttgeaettee gaggtettt agattageaa gaeetetagg
 2820
```

адавадаядя гресядавад гдграстда аддрарссрс агордагада дрададассср

tetttaetgt cacettgaet ggatttagaa tegeettgga gatgtgetta tggtatgtet

teagecteag tigagacteg aatacattig teatgeigt etgagaett taatggitga

240

T80

150

MODOCID: <MO B835644A21>

3900	адссасттаа	כבככבמכבבב	attcctaatc	ctagcctcaa	gtagcccagg	crcractar
3840	Ercactaady	CEGECERCRC	- ಆತ್ರಾಡಿಕ್ಕಾ	crädcrrdcr	atataactaa	tgattggcta
3780	cagatctcac	cacttgaacc	cssggtctct	tatattgagg	cttcaccttt	ctgcattgct
3720	agaegecee	csddcssssc	aacacacatg	ಡಡಿಡಿದ್ದರ್ಧಿಕರ ತ	adgeaageat	racaggecec
3990	ววววะ6วววว	rdardecerr	cccsdddatc	agctatagtt	aatcatctgt	aatggcttac
3600	ddadccacac	tegattecea	ddacccaddr	tttccagag		ವಿದ್ದಿದ್ದಿ ಇದ್ದಿದ್ದರಿಗೆ ದಿದ್ದಿ ಪ್ರಕ್ಷಣಗಳಿಗೆ
0758	ಶತರ್ಧದೆದಿದಿತರ್ಧ	ממשרמשרמה ר	'credroggra	ragrereara	accttgtaac	aagcactaca
3480		Crradorer	ב בבבב בכבב	taagctactg	attacaagga	aadtdctagg
3420		rectectase	attcacacaa	drddccrcdc	•	actcattgta
3360		dcsdsccsds	cccctactct	aattccctaa	cctttaaaaa	aattottaac
3300	หาคาาาคะาะ	בבשרכותושש	בכבשכשכבב ב	catacagtga	tdtdddcsss	gaatcaaggg
3240	ಗಿಳಗಿಳರ್ಳ ಗಿರಿಗ	cgdccccgdg	dccaagagaa	ccagtaagga	cddsssdscs	dddc¢cssds
3180	าวคิดิซดิวอวดิ	radaatataa	garaarrara	aagttacttg	assasstcts	pagagagata
3120	ataagtttt	седдесегад	carcarcrea	ttatacactg	gcaggaatct	agctttggca
	ttagctcatc	garcgaact	dsacccradg	redaccaege	gagttgatat	csddfcfcsd
0908	ggtcaaaga		edesededed	arrererag	taaatgaacc	ctaaagactt
3000	201120112	מחחממממררמ	гдааааааа	эссэвства	tetgtattta	atttgatttg
7670	ttttaagagt		accargrade	Caaattaaaa	unncntggaa	nnnncntnna
7880	_		gagagaga	ರಿತರಿತರಿತರಿತ	ದಿತದಿದಿದಿತದಿತ	ದಿತದಿದಿದಿತದಿದ
2820	teteaagtat	ಕರ್ವತಿರಾಧಿಕ	ದಿದಿಡಿಆರಿದಿಕ		ದಿತದಿದಿತತದಿತದಿ	ಇಡಿಡಿಡಿಇಡಿಡಿಡಿಡಿ
2760			CCCCaaacat	gcgdcgggcg	caccaactgg	ававсадсас
2700	ದಿರಿತರೆದಿರಿತರೆದಿ		tecesesses		cacttcctga	agctaggccc
0 7 97	aaagaaata	5662222122	2242224224	acadadaga	даадъдсада	
7280	tacttctgcc	Cotadtaacc	Sactagadac	agesagecae	agcctggtag	scsdcsdsds
2220	ctcatgtgat		rdagtggcaa		tgggagaat	aacacatgac
2₹60	ರ್ವವಿರಾವಧಿಕ್ಕಾಗ	_	dcccccadr	1111511100	tcaagaacat	ctgagataaa
2400	crdrdaccaa	_	tttgtagttc	_	teteactator	
2340	ctggacttta				atattegggg	ccsssagggc
2280	tgagaccaat					cagagagett
2220	cattcttatg				בקבכבבכבב	
2160		rcraddccda	ctactccagc	transasatt	111211212	acacagactc
STOO	ctttgaaagt	ತಿವಿದ್ದರ್ಥವರ್ಧ	ccractagga	ctcadatctt	agactgacag	•
2040	csdddtddtc			ccaactatoo	ataagggata	aaatgagaag
086T	ctagtgtaaa	ratgragtga			ccctactcag	agagaaactg
1920	aacttggaag					_
098T	tactaactgt					
7800	ಶತದೆದೆದೇಶತದ					
OFLT		gattcaaatg '			_	gatacatcac
7680	ಶಡಿತಡಿತತಡಿದ್ದ	agatagatat i	atagttaatg .		~~	
T620	cagtgcaatc	ccatttatgc o	rtatgtagtt .		_	
09ST	anatagetgt .	garrrrate	tectegaatt	. pepttospt		
0051	ರಶದಶದಶದಶ	390000000	tttaa ctgtt			гдаагдаааа
0777	argacagttt	ceretacted :	rradcrarrr			בבבבממבמב
1380	rdredordos	ccaatgtcct i	gcactcttag .			
1350	catttcaaat	гсааадаагс с	caaagacaga ,	gcatggacc ,		
1260	cs	σεαεαοεαεο ο	crocedrrar	addftactt (tgaaacttgt g	teagettett
7200	ದಿರಿತಕಾರ್ಥ	arreaddeer i	adcractarc :	stacaaggt :	agcatttacc	tgtttacatc :
0711	วาถิดิตาดิตวดี	asgadagar a	מבמכמבכבב	ובשבבשכבבב ו	aaaatactgt g	rtattactot s
1080	วาววดิตตาดิต	cacacacaco e	grandadu (g edopedana	s opoptabati	taaatatata (
1020	ากกากการกร	 6161616126 	adcearacdr :	: dccffffaa :	cagtcaaca t	attadactaa c
096	იიიღინიინმ	ეგიმანის გ	ಚಿತ್ರದ್ವಿ ಚಿತ್ರದ್ವ ಕ್ಷ	idaacacgtt s	adatatgtct s	ttetttaeaa e
006	าดำดำกออาหา	adeadaeada e	gaageceee s	attatagta g	ttttgtaca t	aadaccacad t
078	ใสดาดสกิศสส	iccessaced 6	aardaaaca r	יבבשבבב ו	a titaaattt	attagtagtt t
	וחרררשהרש	gegggdege e	בששככבבכב ו	ובבבמכבמכב ו	g dyddaetaa	atcattaatc a
720 780	ნითიულინინ	arddraarr 6	recessice e	reagtttet :	e tegegatat:	ttatataadt t
	ะดิวาตายาติว	rdagaaaca r	raddreader e	: cccgaaac	aaaanaaacc c	aasassas s
099	กิดกากการว	สลิตราสิติการสาการ	cceredadr s	יבבבבבבשם ו	atabtatabi	accaaqtqaq s
009 050	במרמבמממה.	laggagragr r	caacaacaa c	s sagtttsas.	stectateag a	tettasattg c
0 7 5	ימאמממכבר.	e ppagaaaa	3 Seconocag	cracaddad r	adracrada r	בכבמככבכב כ
087	ירמררוחרמ	n 66marame6	ממכבמבככב כ	משבבבבבב ב	ಕ್ಷಾಂದರಿಗಳು ಕ	aaaataaatc c
₹ 50	วดีดิชดิวดีดิชเ	בבבבכוושש ש	ישמשכבמבשב כ	ctacatcga a	s dassedased:	cttcacactc t
360	Adecedera	- 6168080808	cordressa a	בבשבכבלבב ב	raaaccggca t	ggactcaata a
300	e tonenonn.	, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,		. , , , , , , , , , , , , , , , , , , ,		

MO 66/37644 - LCT/IB98/02133

811

080L	attetgaatg	ccdrraddar	csdsscccss	ccaaagctaa	acctagcatc	caaqctcaqq
7020	ยดากาวดีวาย	ddcccdcccd	recarrerar	arraracs	ccacggccac	atttctaaa
0969	כשכבבכבישר	rccaaaarcc	attcacctca	tctccaaata	gracttcat	tacctaaagt
0069	dccssscdsc	csdccssdsd	dracadddcc	cacctcacag	ccccatgagg	tasaggteet
0789	darcccrcs	credrecede	aattatgotg	atgaattgtg	atgattattg	ctcttttga
0849	стававство	darcasarar	graccaaca	dracaccacc	dattaaaggc	rdsdrdcrdd
0272	ววาววดิวววว	arccaccege	dsactcagaa	gactgacctc	ctgtagacca	ggaactcact
0999	าววาธิววธิธิว	rdraradoco	adddrrrcrc	rrrrdagac	ratttatta	במבבבבמבבב
0099	าาาถีวาาถีวา	פבבבבבבםב	שבשבשבבבב	ddfcaaatat	cttccaatca	actgtattta
0099	ccedrarana	agaccagccc	ಇಡಿ ಇಡಡಿದ್ದ	rrrccdrcag	acttgctatt	cttaattcta
0879	gracharaa	בבמכבכבבש	dradreceed	ataattagct	cagtgctgtt	rrccrddscs
	gactgctcat	าวาธิวธิธิววธิ	reateggrat	дггээдсгас	craracra	tgcactctaa
02 7 9 09 2 9	tgccactgt		дагадагсас		ccfdccfccd	
	ddactcaaac				tatttactct	tctagcnttt
0089	taattotaca		agtgtataca		tagtgggatt	
0779	tatatacaca	במושתשושום	cacacatata	cacacacaca		tatatatata
0819	EDEDETETET	ברשכתרישרש	1016202222	дссевссед	taacccttta	tttctcaaaa
9750		caggettete		rgaaaggrar	дгеадгаагд	ದಿಡಿಇಡಿದಿದ್ದೇಶಇಡಿ
0909				tttgaatctg		aacataaatt
0009	בבבבבבבב	1261111611	ררוששחרששש	שבכברשרשרש	садаравьст	creesdeerr
0765	בבבבבבב			tttgaaagag	scradaarac	ಡೆಡಿಡಿತಿತಂದ್ಯತ
0889	rdddcrrrar		agg-cag-cag	วถาวถึงถ้าวว	rrcccrddad	
2820	ggatggattt		y cacacacy	ממחררוחרת	CEGEGGGEE	dedescede
0972	cagaagaaga			מרררממאאר	ccaaaacgac	ccccagccca
0072	rescararar	atgagtttta				гадасярсяя
0795	tasactacat	ctateceate	raddcaagca	Areandana areanana	aggtacagga	
0855				gageageta	כמבבפבעעם	agaatggcct
2250		aagcttgtga			acacatatgt	
0975	catgtgaccc	ascaddattt	atgttttaaa			tttgatgaaa
0075	_	- eeee	graecases d	Dissension	מחששששרחרש	Caaaaaataa
0789	ttgtatttca		cgaatcatca		22222222	attgtcaaag
2280		attactatat (totttagat i			gatagageta
2220	ברבבבבבם		מבבבככבבכ	1010112210		agtatgagct
0915	gttagtcatg	בכבמשבמבבב		raradaacro		
0015	gtactttata	ctacatttaa	rraacctgtc (6016016111	tetgtettae	
0705	cttacaggga	geddatttt.	ataacttett !	fitagtgatca	ttatatatta	cataatattt
086₹	Liccgctgga	tatecatete	attatgtata	ccattatott ,	. othtettppp	atatatecaa
0267	agagtatgtt	cttaaaatat	rrracaaatc (agctctccat		tecageacae
0987	atttgctaaa	s oforobirs	ರ್ವದಿ ಪ್ರಕ್ಷಣ್ಣದ ನಿರ್ವಹಣ್ಣದ ನಿರ್ದಹಣ್ಣದ ನಿರ್ದಹಣಿದ ನಿರದಹಣ್ಣದ ನಿರ್ದಹಣ್ಣದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣ್ಣದ ನಿರ್ದಹಣ್ಣದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರದಹಣಿದ ನಿರದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರದಹಣಿದ ನಿರದಹಣಿದ ನಿರದಹಣಿದ ನಿರದಹಣಿದ ನಿರದಹಣಿದ ನಿರ್ದಹಣಿದ ನಿರದಹಣಿದ ನಿರದಹಣಿದ ನಿರದಹಣಿದ ನಿರದಹಣಿದ ನಿರದಹಣಿದ ನಿರದಹಣಿದ ನಿರದಹಣಿದ ನಿರದಹಣಿದ ನಿರದಹಿದ ನಿರದಹಣಿದ ನಿರದಹಣಿದ ನಿರದಹಿದ ನಿರದಹಿದ ನಿರದಹಣಿದ ನಿರದಹಿದ ನಿರದಹಿದ ನಿರದಹಿದ ನಿರದಹಿದ ನಿರದಹಣಿದ ನಿರದಹಿದ ನಿರದ ನಿರದ ನಿರದಹಿದ ನಿರದ ನಿರದ ನಿರದಹಿದ ನಿರದ ನಿರದಹಿದ ನಿರದ ನಿರದಹಿದ ನಿರದಹಿದ ನಿರದ ನಿರದಹಿದ ನಿರದ	raarcttata		tactcccctg
008₽	aacacaagac	s dopposes	drrcadcaca d	tracetttt.		
0 を と を	ברכככדקדרק	ascetetage c	ageteacagg	trastatice ;	, 2601222601	gaaacctgg o
0897	cacagtggca	accaddaggg !	acadadtgta g	esectitopi		
4620	reagtacagt	scaaaacaga s	מבבשכשכבב	cactodata (ccccccaa .	aactgctaag 1
095₹	greatttget	agcaacagca g	dascadara e	CCGGGGGGGGG		
00S7	grataagaac	cagtgacaaa g	rcataattet i	actaacttat (raaaqqqttq s	tagtagactc o
0	crrctggcs	daaggcaaac c	caccaccag c	tataadatatt 1	tatootett	tttgttatca
₹38 0	ccttcaggt	tattaatgc o	acacagggc t	, dadopadase	restatador :	taactttata i
43 20	cctgacatc	agtgtgagt o	gcagcagcct t	adagcagta g	מכשמדבמבכב ו	aacacactgt g
4260	gacacgtgtg	gaagtaatta g	sacttgatta g	teatttata	בננמפנננפנ	tacttatttg t
4200	78200	sagatcaact s	saactctgca c	t tecesatat t	ragacraac s	gatcatgatc a
0717	tctaagagt	ctactttat t	ttaaatagt t	ccagtggtc o	gtaaagtgta o	tgtagttctt g
0807	:aactggatc	gcatacatt t	atacacāca t	sacatacata o	s atacatacata	catacataca t
4020	ttacatatg	cetacataa t	tccaaagaa c	tagogatgt o	agettggagg c	מממבשבבב פ
0968	:dayraarrr	aaggtatctt t	rcctggttga a	s objoboobot	ttcaggcat c	ararccaaas r

caceaffera trafferaga gradatecea ergeaggasa etagagtete ceateaatge

tagocagoot gtotgacaca ctggotggco ctcotcacca caatotgaco tcatgotggt

cagactetge atttgaettt gagtegaace acagagtgge ceagagtgae tetetgette

agettgatag ggattttgat ateagatgta ttateetggg teatgttegt gagaetgget

гаааддадст дааасасдтт тетсеседтд гдасасастд даааастдаг даадсесаа

ರ್ಥರ್ಭವಿಕ್ಕ ಅವರಿಕಾ ಕಾರ್ವವಿಕ್ಕಾ ಕಾರ್ವಾ ಕಾರ್ವಾ ಕಾರ್ವಾ ಕಾರ್ವಾ ಕಾರ್ವಾ ಕಾರ್ವಾ ಕಾರ್ವಿ ಕಾರ್ವಾ ಕಾರ್ವಿ ಕಾರ್ವಿ ಕಾರ್ವಿ ಕಾರ್ವಿ ಕಾರ್ವಿ ಕಾರ್ವಿ ಕಾರ್ವಿ ಕಾರ್ವಿ ಕಾರ್ಲಿ ಕಾರ್ಲ ಕಾರ್ಲಿ ಕಾರ್ಲ ಕಾರ್ಲಿ ಕ

caageteagg acetageate ceaaagetaa cagaacecaa cegttaggat attetgaatg

0954

00SL

0**77**L

7380

7320

7260

7200

OFTL

3/2000 SAO 8832644A2] >

17550	tactagcatg	attacatata	ತ್ಕಡಿಡಿಂತಂಡಿಡ್ಕ	acttcactta	tacgaaatgt	gctctatgta
09111	acagcagaaa	acttctgta	atattcacct	cactdatgaa	taaddaatdd	2016261126
OOTTT	าธิวาววธิชาว	csaaagrgca	graatacaat	gcattctaat	tataaatcaa	agecagtttt
TTOTO	atggaactca	aattcatttc	сстаддіссва	ccattctaat	atttagaact	ttcagaaatt
10980	ttagtgtcat	agaattgtgt	acaatatggc	aacacattgt	tatttccaca	ctaagaatgt
10920	dsraccsara	tgattttgta	ассаддассс	ttcagaattc	בנפכננמננר	tgtagagcat
0980T	tactatagga	gaatatttga	dscrradcr	agaagaactt	aggtttgtat	tgagatggat
T0800	ctacatactg	ಶತತತಡಿದ್ದರ್ಶ	таставадат,	ttttagcagt	atacatagaa	ttattttaa
OFLOT	cdsdrdsdsr	atgctggtct	tctgtagacc	tggaactcac		ctgtgtagcc
T0680	csdddrrrcr	ttttcgaaa	ααεεααεεαα	בבממבבממבב	מדררממדרמד	בבתתתתתקבב
10620	בבבבבפבבב	attttttgt	адассссас	aaqaqtqqaq	actdataata	tcaaacactg
J0260	aacaatataa	ggcatataag	catgacctca	agaagtttg	atttcaccat	atttcttgt
0050T	acacttcctt	gcttctcagg	caaatattct	creagatett	gccatttatt	gaattottot
07 7 0T	tcacatggaa	cattttctgt	cctgatcttt	acadaccttt	ctasatccta	
10380	tcatatcctt	gcatccctgt	cacttcaaca	accerdadea	cattttcatg	1011011160
T03 50	taattaactc	Etgteteete	aaattagata	ttattaattt	taagagtttg	gaacctaasta
10560	craracrarc	gattagaggc	ಕ್ಷಾತ್ರಕ್ಕುತ್ತ	reserected	רככררכרמכר	בנכננננגנ
10200	ddraddccrr	aagttacctg	trgtctcact	rrracascs	בלבלבלבלב	
TOTTO	ರರ್ಧಾರ್ಯ	gcaccccc	tcactdagct	atatactect	cdfdaffaga	בבמבכפבמב
70080		rcargctagg		gccccatccc	treeteeeee	atagctgtgg
10050		teceagactt		cattttaaca		craggatatt
0966	gaattatgog	aaatggtcaa		atgcattatg		CtCtaaaaca
0066	agataaaatt	acgtgtaaag		trctagaad		tagataactt
0₱86		ragergrett		tragattatat		descacata description of the second of the s
0876		tctatagaaa		cagggttcat		dddfcaccat
9720		atttaggcag		tcacagttac actgcattaa		gaattgtaac
0996	tratoottoa	casastastt		tgggtcacaa		
0096	ortacattat	ttatcagata		catttatac		radcrrrrsc
0756	aaactctota	scradccadc		רממחכררמממ	rattgttcca	
0876		tcaggaaag		הפשמטרה מרשרשרה	S SpeSennin	ctcttaatag
9420		taactgctta		tatettggat :	าากะวะอดิยว	darrardara
0986		tttaaaaata	339292929 339292929	, 10001100116	radcreary	ccaccattcc
9300	ttaaatttaa	gagattaaag actaaaattt		creageore	garrrgcorg	aactcacaga
9240	rerddeerrd	_	משפררשכפש משפררשכפש	adcerdacer i	atataacccc	aggtettace.
0816		מבבבבבבב		מבשבבכבב כ	даатстатат и	gnacagtatt
9750	ב מרממט רמטמ		greatctegg a	raarccago i	כבבמכמבמבב	aatacattcc
0906	argradutan	scagatatcc a	3363633335	rerracess s	aattaaatta i	ctgacaacca
0006 0768	216262222	radarcaceg (accecaedde ,	acragaraa g	catggcttat g	ttactgaaaa
0888	teretteaag	ເວລາວາາ6ລວລ	מבבבשבבבם מ	ಗಡಿಡಿತತಡಿದ್ದತ್ತ	rdracagacg s	ಡಿತಡಿತ್ವಾರಿತ್ವ ,
0288	1661100600	בכשרששכשכר י	racacarc	taaaccctt s	attcatgggc a	ccadtdctdd a
0948	2622266667	6612612261	, ದಿರವಿಗೆ ಬಿಡಿಗೆ ಬಿ	araecepte s	ctddcatatg g	adracadtac o
0078	וווווממממוורר	uuuuucucuu i	: euuuuuuuu	i gasattopsi	rtattdatag s) Telepiteeni
0798	~~666~~66~	_ ინითინდნი:	cceddacccr	iddcydydcy (rtetaccaat c	teratorott (
0858	าวด้าวดิวดิวดิวดิ	coedcecar :	:rcrdddgdg (iddefetdeg <i>i</i>	rtaaaqctga g	atamatmost c
8220	ชาวากิกาชาช	cdgrqdqdr v	serrecede e	sattootte:	ctttaaaatt c	cactattede c
09178	ppกกกกกกกก	1511088101	crederaer :	regretati	rtatectget g	ettttacttg o
00 1 8	กาวกากควา	เลอาดียววออยว	ideedscoee i	icceatgaag s	ctddcctdgr 6	tetatteact o
8340	רכשרורורשר	ากกาวอดีตอาว	r derderdd r	:drrcrdddy c	tacaqaaag t	assatatase t
8280	cedrorord	idoffaagaa i	iddfaaagag s	gacatattt s	tecetadada t	ataastttaa t
8770	lcgacccaag	rtcaaacca g	frccagagg 5	aattccagg c	actttgcat g	gctcatggaa g
0918	aataaaact	sagtettete t	ಡಿತತಡಿಡಿಂತ್ರಡಿ ಕ	cgtgatgga a	tattagtet a	ttagaattgc s
8100	್ದ ದೆ ಇ ಆ ಆ ಆ ಆ ಆ	ıgaggaatga c	gaacatacc s	cacaagttg g		tecasatgee t
0708	satgetgeat	: ರತಕರರರಕವರ	್ವರಾತ್ತಿ ಕ್ಷಾಗ್ರಹ್ಮ	acagacaca a	tacaacaca c	acaagcaact g
0864	cccdcccag	caagaacte t	gcatgttaa g	cagageett t	asasadtac c	tddcsacttg s
7920	בבבבמבבבב	וכפרבכבנקב ב	.darrrasas s	cagcetta a	taataacaa c	tectetaggt o
0984	:csggtgacc	actetgeet c	e ddodddeos.	g treedada d	e dioditioni	getaageeag t
0087	:ಇಇರ್ವಿಇವರ	ataacgtgg c	agatetaaa q	בבבבבבב ב	o ogazaceci	acagcaggta a
O₹ <i>LL</i>	: ರದ್ಧರತ್ರದ್ವತ	. ಕರ್ವಕ್ಷಕ್ಷಣ ೧	geagaette e	tteetttaa a	p sacstassa;	aggetgtgg t aggetgggee a
· 089 <i>L</i>	gctgaaggc	taacadddc q	darcreges a	ctetectet a	n ttntnatny	aggtgggacc a
7620	tacaaqqac	actatabte c	cettaatda	1 117755155	e nn4e44440	כבבבכבבכ פ

7₹880	aaggtttctc	ತಡಿಡಿಕಡಿತ್ಕಡಿಕ	agattgctgc	ccdtdaggta	tgtcctctgg	tgagataagc
74820	ααααεεεαε	dsactggttc	tddagcaatg	ttaccccatg	tgaaaatata	taagtcattt
094 7 T	gttaagttct	aggttaagaa	tecteaagge	cegetittat		######################################
00LTT	atgtgaacgc			ataaagaaat	gaatttaatg	acgaagtgcc
0797T	tttatgtaaa	cstddsddss			· ·	aaagaataat
08S₹T	attagttatt	aactacggag	בבשככבבשבד :	aggagtacag	• • • • • •	בבבבמבבב
1 4220	rdsaatgtag	dcttdagttc	actaactgtc	atatttaaat		tattgtatat
097T	rrcaggtaca		ddcfcfdfag	gtttacttg	ccatgatgca	tgtgtattt
0077T	сваваттава		gaaaatattt		tgaactaaat	aaaaagcttt
07E7T		aatgagttag		grassrasca	aattgacttg	ggatacttt
7 4 280	gatatttaa	rdddadrara	tetecagtee	crasaccrrc	creatatera	tctgagaaga
1 4220	atcaccagtc	cradascrfa		caarracas	дадстасдад	tgagtgtttt geetaeaaet
09T#T	acattaggtg	tcagaaaaac	cccstggagg	9ffccfdddd	gratoratgo	ccrccrddcc
00T#T	tttatgtgtt		atttattatt	tetttaegt	recrangace	tattttaatt
0 7 070	tetateette	בבבבבבב	gacttttgtc		attttatgta ccttaaagga	1156111101
13980	tectagteca	tgtttctggt		actgaatgcc	addoccotdad	agttctttaa
13920	ttatgtatta	trrigcocat		tgtagacatt	trtagtccat	ರ್ಷದರ್ಧ್ವಕ್ಷ ದಿತ್ತದ್ದಿದ್ದಿದ್ದಿ
13860		гдгэдсгдэд	caccaatctc	cactgtggaa	catacacagt	accetgeese
13800		caatgtaaga		agtgccctta	ddccrcadcr	ccaccagcag
73740	tttcttaaag			agccggcctta	•	agcettetee
13680	ממרמממרמר			tgctgtctcc		darcedaera
73620	ratagactgc	ctctttggat	2622226265	במבבבבבב		dagtgtgaag
73260	greceesa		gacagtttta	1211111111	קקקטרקטרק	аассдассса
00587	rcraagcrca	~	acaddadact Tagaddagact	משמחרומרממ	agreeaacag	aatacctagg
13440	aatgttcgga		cgtatattta		tgictagatg	rdarddrdrd
13380	cagaaataca			tecaaaetga		аасссдедад
73350	tcacggttcc			ctgtttacta		
13260	adarcerear		_			cddccsdcds
13200	raggitacaa			agacagtoto		
13140	ccaggctagc				dragadreas	gradaracar
13080	gracacac acttttagg					tagtatttt
13050		tteetttaag				grcatatgtg
175960	gaacatggg				gatttttg :	aactatatct
15800 15840	satttaatga		בבבבבבפם		adraaacttt :	actttgctca
12940	razababapa	מכבככמכבמ	agttaaaatg g	tgaaagaca s		taggacatgt
12720		adcraceds o		cadurearea c	tacagttgac	adarcraccr .
15250		gettgtgag g		್ಡಿರಿಡಿಕಡಿಡಿಕ್ಕ್ ಚ	בבבבקבבככ ו	rdsagatgag
17660	gergraced	racrecese 6	ctctcaggac s	satagttaga i	setetgeete (αςαςςτεςτα ο
75240	adaccede :	adderdacr c	stetgtacae c	cradascrra c	cctagtgttc (gctatgtggc o
12480	credarcrr	receptada	aaactaaaag c	attttaaaa s	atcatacttt :	aagcattta :
12420	αςτεαττε	craradrea c	tttaatgtg t	gadaratatt i	rffddddad 6	
15360	22225226	3 กาศกากคำกำ	r aaadddeea	crecrager r	aaaacacaca c	999999999
12300	ומממממרוזכמ	:48488444 c	saddoraca c	ಂತ್ರದ್ವತ್ತಾರ್ಥ ರ	cacatagtt o	accederate of
15540	gagttcaag	igtggatcac t	agcagaggta a	s popiastics	cagagggac c	ttttaagaaa 1
15180	ಗ್ವಡಿಕ್ಕರೆಗಳ	atcettgaa s	יבבַבמֿבָמֿבבַב	satttigtac c	radccccatd t	aaaatctctt o
75750	೧೮೦೪೦೯ರಿತ	gaatootta s	agcagcaac t	ccrcracas s	seemeeama:	ttatgggtac c
15060	:ಇಡಿಇಡಿಂಭಡಿಡ	lagttagagg c	rttggaactg g	o potatatec e	. e====================================	cycffddddc f
75000	ιραοορααρα	actgtgcat g	atgeetgeg c	coctacata t	s eesseseee. Tatattaaa	reeccedad t
07611	:במבבכבבכ	tttaaacat t	ittttgtage t	ittacttadt s	gaararaa .aaatataa	מבבבממממ ב
11880	tcaacagtt	taaatatta t	tacctotaa t	: 22222222 : 222122632:	. Jeansangs. Thistesen	cetttaaaa s
11820	101760011	t tastattoo.	וככפבבבבבי	o dinascezo:	y jasastane.	atgggatatt t
09477		tttgtcaga c	rafgaracc r	s etthythoped:	2 424546299 2 4245462291	gageeaegae e
00111		dayyagacy a	n eassacee.	o paspaces o paspataspi	ונמנמנכנטט מ	cctatcatat s gagccacgat g
07911	-	tateaataa a	a sanasanga addatanga	e neenetont.	1 226266262	taaaaaccta a
08511	בפרפרפר	a eercerre acttattta a	i itelitebe.		6 662226222	cacacgagac c
17250		a pasectora atateceada a	T DEDEEDTIE	- ~==+~++=+	. caaaagcrg c	acccacaget t
09711	C C	s perpetage	4 246462642	רובשונכשם ש	יבבבששששם ב	crcartgctr a
00011		u asestosoo T seestosoo	- ~++++++ - 61p161p1p	בנשבבנשבב כ	gcagagaca c	aacsacaara a
07811		4 567111100 1 567111100	aaacagaag a	racerece e	gcacgcatg t	rdcsdrdsds s
77280	. 6626227430				•	

3/12/2001D: <MO 8835644V5 | >

	66	-666-6	66nna6anaa	ทาวาทาทาที	tactttgttg	reardacecs
7824O	defacedccd	atagasasa	ppeepptent	5121152512	rraaggtetg	22222222
18480			pseppetest	2246662442	radacreada	ררששההההשברים
78420	ttctgtgtta	tetaetaget		rgarggrerr		ttatctagtc
18360	ctacaaagtc	cacttcaaaa				
18300		tagccatcct		retetadece	gcatatttg	petentoses
785₹0			craggaacca	Coccaccar	cccscsdcsd	11060601111
18180	tgtcttaaag	tdatgtcctg	Egecttccag	tagaatteet	acceatecea	assectesdo
18750	agtagtctat	dcaggaagct	£ಡಿಡಿಡಿತ್ವರಚಿಡ	tetgtagata	ageteagate	ataddctdct
J8060	сಶದಿಂತತದಿತತಂ	acaggtgcac	gcatagcagc	gactaagaag	agagcaccat	
18000	gtatcttgat	ರಿತತರಿತತರಿತ	caacttaaag	tdactggaag	tggtataacc	ctttgtcat
0764T	racrarere	rrrdcaradd	ааадаасаса	ಶತರತತತತತ	ttctctddat	ttaaggegte
08871	асаасддасс	rcrascucra	двасвасвая	ಶತತಿದಿತಿದ್ದು	Cacttcttt	gataaataa
17820	ggaatttaa	tttccttcct		tecetateat	cttccacaat	τεοςεαεςεα
	gatgtcatgg	cactgttcct		ಶತರಿಂದರಿತ	raccracrcs	
09771		gccccacaca		ತ ಡಿಡಿಡಿರ್ದ್ಗಡ	сасавосида	
0077		ttaggaagta		gttgtactga	ctctcatgta	
079LT		_		tagtttggaa	ccatgtgcgg	catcacaaga
085LT	cagatagntt			tcagacacaa	ttcacctgta	_
77520	gectaggtee				dddccgrgdc	
09 7 LT		ccatgcacaa		ccatgctcag		
00 ⊅ LT	cacctgactt		ctgcaccagt	· _		
17340	gagtaaccct	decetateed	_		gacagatgct	tagcctgaat
17280	tcagcagttc	tecttetaaa	actccatgct	aacatcaccc		attcagattg
17220	ttacacctgg	gacgtgaatc	atgccagtca	tgetetgeee		ccactctatg
09141	ರ್ಧರವಿಧ್ಯಕ್ಷ	agataatgta	ragetggaga		αοςεεαοεεα	ctataaagag
00141	racadcacac	_	ದಿದಿಡಿತಿತಿಕ್ಕು	gtttacagta	ggtcatttcc	raccccrara
07071	асдедорова		_	cccttacgcc	tagtcccata	ತತತಡಿಡಿದ್ದು
08691	rccrcddcad	dscraccrac	rccaccraag		<i>catctgagg</i>	cagatactgc
	aagcacacac	מבככבבבבכ	arcaaactga	caraccarcc	tecesasas	ccradarrcc
T 6 9 T 0	agctgggtga	מרביידידים	gecearegg	ccdcdadcad	ctactgactc	grtgrtaact
09891	_		רכשרחרכשים	ggcggccgcg	ссссявдска	crcacaaata
T6800	ttccataagt				cccccagas	recategrea
07L9T	cattaggtgt	ccactcccac			craccarges	
08991	בבבמבבבב	, 1611216161	1611216116	ברבשבכככבר		
16620	במבבממבבבמ	1 4440044440	ברכברמממבר	112221211	caaagtattu	
0959T	cradacarar	readacaada (rradcaaras			
00591	_	בבקכבפבבם				
0 55 9T	sedgedodda				_	_
08E9T		taatacgtt (_
T 2 3 5 0		gaccctagaa o				_
16260		agatacaaag o			aaatcaaaaa a	
16200	tctaatttt	gaatagtttg t	aacactgaat g		caagaatgac i	
07191	tgacacaaa	cgraaagtac t	atgcattctg t	agaaacttgt :	attatgaaag :	
16080		במכבבבמבמ ו	edreattgee c	೨೦೨೦೨೦೩	s tttotottt s	_
T 6 0 2 0	tccccccaga	cttgctcaca g	caagtacact c		tettaccetg (
09651	206262200	agaregerea i	ರ್ಧ್ಯಕ್ಷತ್ತು ಕ್ಷಣ್ಣಿ ಕ	adrdrcccca i	tacagccatg :	gattctgcag
00651	тесасасся	явдассссвя я	agctatatac a	screatacte s	agaacttcag a	cagaagtgtg :
07851	gggcgggdg	agcacteag g	מבתבתבבכם	eddcatatg c	gcctgttata g	ttgaatacca g
08721	רממממממממ	בששבבבשרה ו	מבשבששבשם	מבשמבבבב ו	eagctttaa t	בבבמסכבבבב י
	מששרר הרש	610181110	geaceage of	משכשבשבשש פ	caatggtcat a	caatgtggag '
72720	מרורמרומ	1 161111618	a noneebabab	בבבבבבבפב	בכבשמבמבכב ב	כבקבמבבכב ו
12660	222262776	1 1011010101	5 454552424- 5 6565778556	מרכרורשכש ה	าวดิสสติสติด	ggcatatgaa
00951	112122126	e eccasasas.	, paperates.	י ברבקה הואים בי	cattctagt t	agrecedges o
12240	agateteaa	s offetttoer	etenethno.	, nananakak	, 100101100 1 1660110161	racraracec ,
08ħST	rtttaataa	ittitatatai	carritage c	1 621111010.	grattaggt t	6662226622
12420	:מַבְבַבַבַם	icadctacad t	dagaceacs c	* Prepettet;	* 660080011t	rcddccrddd c
72360	sccagggcc (iggcaacagt c	ctacacacc a	CCCCCAAA9 C	secresced c	, 212220tr21
12300	וכככבבבכבכ	s dododdogda	o fototobjo;	בכבבמבכבב ב	σερτασασαα σ	rerdaraaa
12540	:ಇಡಿಡಿಂತ ಕಡಿಡಿಂ	ctgtataga c	regacecete t	ס ססססדסטססן	ctaccaact c	aatgetgaet g
08151	tettteaag	tcaaatctt c	gaaaaagaa t	tgtaaaggg c	aactgtttt c	בפרברבפכפ פ
12150	ccdfdfcag	sagagtcat t	retectagt g	ಡಿತದೀತತವಿತ ಕ	cacaaaccc a	tttaaaacaa a
09051	בבפבבבכב	laaattaaag t	tttattgta g	aatttggat t	s ipipiojiji	aaattgcttt o
00057	್ವವಾರ್ಥಿಕಿಕ್ಕಾರ್ ಚಿತ್ರವಾಗಿ ಬಿತ್ತಾರ್	יבבכשמבבכב ש	cecatgag a	gagttagtc t	laagccccaa a	ggagctcact g
0767T	~~ ~ ~	dccffagfg f	ನ್ವಿ ಕ್ರಾಕ್ಷಣ್ಣ ಕ್ರಿ	c <i>c</i> cdddacd a	ircagcatgg g	ctgagtaaca g

22200	cstdtttdgg	ttctttataa	dctttgagct	בבבבמבבבבב	cttttatgct	ccactctat
22140	caccaatcac	ctacacccag	rradacrecs	caagetggee	ctatgtagac	caaaggcccg
22080	ρετααετααε	actatgaagc	agaggttcc	ggttttcagt	ggrecagtet	tacttaactt
22020	rrarrcacrr	aaactcttgt	caggtaagat	gtacatggta	аадстдаата	atttccatag
27960	araagtccat	tccatgactg	aaatccacca	gaaaatactc	aaaggttcag	aggasatasa
27900	rddrrrarda	gatgtcacag	Egcaatttat	gtcatttaga	tctatgaaga	tacttttaat
SI840	ccscccscdr	ಕರಿತಕಿರಿತಿಕಿ	sctgacacca	tasascacgt	cttgcagtat	
27780	aatctacctg	daccacateg	tectaattat	tttcatgtat	agatgttgac	
27720	rddradrrda	carctattg	ctttaatttc	tggcattaaa	tgtgaggtta	acattctaga
21660	atgtatgtga	tgagatggca	ttgacctaag	gtcaatgggg	atggatatgt	radatottta
ST00	ctcattcatg	cdcagcatat	cccagcttct	gaccagcctt	tetteceaag	radarceres
572 ₹0	aggetegeag	atcaaagaga	greagaactt	tgtctttctt	ברכברתבם	oractactac
21480	gcttagcaag	ಶತರಿತಂಗಿದ್ದರ	tgtcctaaga	tgctgagttc	gradaddaa	dccdggcgdg
SI 450	ತ ಡಿತಡಿಡಿತತ	ddadcccadd	agtacatcct	tgaaattgta	gractcctg	авадаесств
27360	atagcacatc	drcdcrdrrc	cttctcccta	cttcaacccg	gcagtgcaga	craasasas
27300	decaetttee	tecttatgte	tagetgttet	caccatcctg	tagacagtcc	aattotteg
SIS #0	gcattttaa	ನಿರವಿರಂಭರಿಗಳ	tttctttccc	taggaatagc	tatgtggttg	tgaactcaga
27780	cccffdddsc			caagtttgtg	agacttcata	ttgacttttc
27750	cccagcttgg	agaaaacttt	ttagctgaca	ggcgaatgac		dagacaagat
27060	taaaggtcct		agtttcacag	dasaccraas		aagtctggtc
ST000	gaaaaataaa	ತರಿತರಿತಿತಿಕಿ		ddccsdcrrd	ataaatttga	radasaasrc
20940		ccsgctcctg		ggtgttgcaa	aaccagaaat	atactggttg
20880		grccraataa			accacatcta	atatattaac
20820	tattccctc	gctgtagatt		gtacagacag		saraascraa
09202	ddcagacatt	ggttatacta		gttataggca		cecdeecddd
20700	ttcgagcaaa	ggaaatttcc				acgtgaagac
50640	tattcatgcc	drcagggrca	cctgcactta	tcagaaccga		atctagtgcc
20280	grtrcaggac	atggcttgaa	ರತಿರ್ಶರಕಿಂತ	tgggaagaat		cagcaagtat
20220	agtatacacc	aagtccttca	cgttttcata	ttgtggtcat		teceataaaa
70₹00		attattagtt			סבככבבבב	
20400		tcactgtgtg			radadttact	ccssgasggc
20340	ddfedeseds	aatgctgaga				cddddcccsc
20280	•	tgtaataaga		дсвавсасад		gatgtatatg
20220	,	aacataaaca	ctggtgttat		ವಿಡಿಶವಿಧವಿಕ್ಕಾರ. - ಪ್ರಕ್ಷಾನ್ಕರ	
20160	ccccaagaaa				CCALGCTGLG	
00102.		ragagacgca ;				
200₫0	cgatcttcEa	_		tttcttgtaa		greeracea
0866T	ttccttcaga					agatttaact
77950		ctgctcagcg (tecagecagt		aaagtttatt gcaacataca
09867	aaggtataat		gragatttcc	ee-	tatgeceaca i	geedatttatt
00861	Josesasase	caatggactc a	araaraccaa	ractagatta		gccaataaca
0 7 26T				. Delepteest	atccatatta :	crcaaagtga c
0896T	ggttactctg	בכבמכשמבבב מ	raattcaatc	Haarcaace :	sattetennie	רמרררממחמ ו
7 9 650	radadotata	erenesezza Podesetesse	titopotier		aacagttta i	ורורשרושרם
09567	actattagea	resarceara :	tractasttt	. gamammage	ಡಿಡಿಡಿ ಇದ್ದ ಪ್ರಕ್ರಿಸ್ತರ್ ಜನ್ನ ಇದ್ದ ಪ್ರಕ್ರಿಸ್ತರ್	במחשר רבבשר
00 5 6T	aaaaatcatd	S Defeatified	s sestestop:	bieiesiinr	, 4545464564	ggctgaagag cagec
0776T	adacatatad	S DDIEDILE	: compensor:	, peetettete	רכשששנשארה :	tcataagttt
19380	totctataat	reparament	, 125105211°	. sestantese	2 11De1eee1	acactaggg a
19320	artactagat	trassaceca :	111106010	totecontee	: בבקבקקטקט מרושחרושר :	ctatattett
19261	natotaaaat	reseatesat c	. gadappaasi	. eteeptepte	מרמשמשמט ב	caggagttaa
19200	elepedelle	s remembers and	s sitestini	, 166716161	, วะยะยะนะนะ	cagttctcaa g
07161	2444226642	tttattttt i Tattdacatt i	t sentinunte	. 1e11e11n1.	, 44%ebbe44k	agtttagtt :
19080	tostasosta	1 111111611	, tentetness			aagigacata g
19020	restrasta	t totaliator	: 822822222 : 2111012211	. Topothente	. 610640041	actgtgttt a
09681	reacetanet	, agasassass Tanananan	. UULBLELLI ATANINETER	. tenennein.	, <u></u>	atattetett
18900	Desessor	t intiteesse	. 201616311-	, 545555555 , 545555555	geagacaaa .	ataccagett i
0788T	127222726	: 150110011;	. 612924440.	daccigadad .	. ಇರತ್ತುತ್ತಾಗಿ	agtttaaatc o
08781	1112622211	sagtgetgge c	phantante:	agendadar o	saatgeeace	cctcccaaca t
78720 78660	200000000	יישרוכבששש ו	rraageeec a	יכבבכבבבפש פ	במכבבממבמ פ	caaggcccac
00981	מפרבים במני	dacaggasa r	secedaded o	מבבממכבבש כ	ಗಡಿತ್ತದಿರುತ್ತ ಬ	øddccøfccd s
ししつひし		,				

MO 66/37644 PCL/IB68/07133

aattaaataa ggagttttct tgttgttgtt tttttgttt tgttttgttc 25860 cataagcaag tagcaggctg cagtcacagt ctcttattga tggctacaca ttgtatcaca

tgtacatggg cacctggttg tatggaaccc tccttggctg cctgtggttt gttattaaag

cagtgttgat cttggggagt ttgactgcgg tcatgctgat gacggagtcc ggaaggaaac gaaacaaatt teetgggaaa agtgtteatt ceagactaag tgtgaagaag aetttacett

gttttccttc tggccatagg ttgctcatag agttctatga ttcaccagat ccagaaagaa

tazatggttt atgtttgttt caacaccaaa aatgtccaaa actgaaagat caattctgtt

totgoactea tecaagtett etaaggaget gtacttgete ageaagtact caatacetaa

ggatataata teatgteesa aetgtaagga gtgaatgeee teeegtgeet eteggacaee

agtoctcatt aaatatotta gatocoogco gigttttgat ttttgtttoo cacgtggtga

ctagccaaca cccatctaca taaaagtata taaacatatc tttatcttaa aaatccccga

בקמכמככנכ בוכנקקנכנכ בקבכמבמב מקמכמכנ בקמקבמבמכם קמכמכמכמכמ

משלנכמלנים לכמשלממכה כמלששלשטני לנשלממבל ממכונכמלמל נכממלשהכככ

ತಡಾರವನವನಿಕ್ಕು ಅನಕ್ಕರಕ್ಷುತ್ತದೆ ಅನಕ್ಕರ ಕರ್ಮಕ್ಕರ ಕರ್ಮಕ್ಷಕ್ಷಣ ಕರ್ಷಕ್ಷಣಗಳು

tagaaggeet ttegttttat gagtegggtt ggtggaacea ettaeagatg gaagatttae

ttgageteca geeteageeg tttteattgg ettatgggat gegageetg ggagaggage

cagateacet caggaggaaa actectatge tataagaatt tetttettt egeatetttg

agteaceaca tecagegact aageatteag atatgtgaac etgteggage etttettaet

tetecegace cetatacaea ettecetecaa caaggetaca cettetaate ettettaaag

agagacagac tgagcctggc atggggttett ggaaccteaa agccteteat ecetacecea

ತಂತರೆದೆತಡೆದೆಂ ತಂಗಂತರೆದೆಂತರೆ ರತರಿತತರೆಗಳರ್ವೆ ಗರಿತರಿತರೆಂಂತಂ ತರ್ಗಂಗರಿತಗಂಗ ತಂತರೆದೆಂತರಿತರೆ

catttagttg gggaattcct tacagcttca gaggttgagt tcgttttcat catgctgagg

ctgtgctttt gtgcctgtcc tgctgttaga atcttacaga ggaggatgaa tgaatgaccc

tgtatcggta acteactgtt gtagetetgt ggaagegget caeaggeagt agggaegett

aactiticaa cecageeget caggattatt tigatgatgg gaacaatgta agaaggeeta

ttgaaatgaa gotttatatt agatttatgo ottgtgttgt cacgtgttto tacctgacat

tttcataat gaaattgtcc cctttcttga gttagtagaa agtattacaa ggatagagg

ctgctgtgtg ataagagcet ttectettea getacaeggg ggacaegagg etttggggtt

atgagteact ataaaaatea tgacatggtg geetacetge agtgtttget ggacagtagg

cttttgtga actogggtta aatcttatto tatcotttog tgiicacatt gtacatttto

ttattccaaa gtcgatagca cagcaaagt gaaactaaag tctgtattgt ttcaagaatg

catectacat cagtttgcct gttgatttct gtaccatgac aactcaacac agcgatgcgt

catatactta aatgtagcat gtttcatggt gcgttaccot tgtttaacaa ttaagtttaa

aggagictit ttatcticat tatttagtaa atactaatca tacctgcata gacaagacca

agagetecag catttagaaa gtgcagttea accaaatttt acteteagat cetgettgaa

саадаасаса ЕдааааадЕд дсеесаедад сдсееедада саааадагад деаадеддеа

cagtgcccaa aacttcatat tcactttgat cgtatagaca gaaatgaagt tccagaggaa

caaattotaa actagtaaga cgtgaaattt tottottott tgrtagagtt tototgcaaa

tgittiaaat tottitagic ttaatgitto attittacca taagitacti tgiataatca

atgrectaat gaggttttaa tttcagctta atactgcaaa tcataagtgc atagctttat

ggatggctac ttctccaagg cttgctgtta gaagtcagtg acatgggctt aacaagagat

ggttttctgt tagtacgcag agtgagaggt ttcttactga tgtctgcgta cctagaggaa

gacagtaaat gaggaaatgt aaaatgtaaa agattctaat ttttattt ttaaaggtga

gcagattgat attcaaaccc agccagtttt cttaaatact ttgtggatgg gattggcttt

בקבתונקשקי ארנככככנפק בקכנכבננ בתבנוקבת בהתבכננכם קשמקקקקה

agtgtgatga agacttgaag tttagggaca ttttccctcc ctggccccac tcaccccatc

gaatattatg teacactgaa catgggatgg aagacatgtt etgaggaatg tetgcaetee

cagtaaatac tetecagaga ttteagatga gattetgett cetggtaaac aggaggecaa

agggtgggca aaatcgaaca cttactcttg gagactccct ttatgaatat taccacactc

agtigtigte tgecetaeae aaggeetgga tttagetece agtageaeg aagggaggeg

tyaacaatot aytaaaatoo taaytoayya aytoayyot yayytyoayo toaytayyay

tetettataa tegeagtaet aaggaggaag aageagaaga tgatgaetae agggeeagge

atgagagtag taaggaagag agaagagaga gacgtggtat tttgctgcag actaaagaga

ತಡಿರುಗಿತ್ತು ಇಡುತ್ತು ಕ್ಷಣ್ಣ ಕ್ಷಣ್ಣ

ctttgagtga gaatagttca ggtaactata gccacagact caacatttga acatgggaac

aagactcaga actottgoot ttgtcagtga caaagtgaga atggotgtga agtgacgtgg

aaggacattg tcattattta caagaagaa tatggtottt tcccaacatg ctagaattta

aaagettgtt tttgtgtgat tagateetgg eeteacaat geteggeat eattttactg 24900

tgtcactggc ctcttgttgt gaagagacac cttgagcaaa gcaactcttc tgagagaaag 24480 tttttatttc tcttgtctgc ttttctaatt ttatgggaat aagaacttt ggtaggtctc 24420

caagactgaa gcacgggtga gcacaacac tttgtgttgt gggaaggaag ggaattgttc 24060

00897

071.57

52620

22260

22200

52440

25380

25320

22260

25200

S2740

22080

22020

09677

24840

24780

24720

09977

24600

07577

24360

2₫300

24240

24180

54150

24000

23940

23880

23820

23760

23700

23640

23280

23220

23460

53400

23340

23280

73770

33760

SSTOO

23040

08677

22920

22860

22800

051.77

22680

22620

22560

22500

22440

22380

22320

22260

INSDOCID: <MO 883SQ44VS I >

29520	ractettage	дассссдаса	ttttaaaaaa	attaaaattt	aaaacgaact	gcataccatt
79460	agaccagcar	cdcsddcsss	£ddacacaca	caccttcata	gggtatccaa	tatagtteca
29400	ccdaagcaac	ಇರಿಡಿದ್ದ ಇದ್ದ ಇತ	cagcatccac	gttccattcc	cagagaatgg	tttgcatttt
29340	d 9dff3dc9d	888828888	rraargggca	aggatttcat	catttttgtt	ttagttacaa
	racredadge	racagagaar	ададаадсся	сясясяяяся	tcacgccacc	aagatottot
79280	202242224	REESCOOP A	กิดวาวอดิตตว	десевенад	agtcagcctg	grarrrcca
29220	gagaaagaga	222266222	שרבהאאממממ	2668222822	cscstggctg	ಡಿಡಾರ್ಡಿಡಿಕ
29160	dedesesses	1111006001	222222422	שפררממממר	tgactgactg	agrarregeg
29100	ataacaacta	ppisoistei	6211611212	1665611055	agactaatat	6166111081
29040	saatsasats	astatcatat	titentenes	ttggaatatc		
28980	tasasatcaa	agtattagta	ettteneste	teteenesen	ttacattata	16111116
78920	tattaccagt	taacttaaaa	asacattorc	1260661121	בבבבבבב	5225254445C
78860	tagecagget	gacatgtcca	acaasscarr	ctasatattt	agattaaaga	Diinipa
28800	caattgtttc	ttaacctcaa	taaataaact	tataaaacac	taaggtcata	Senanzaza
28740	agaaatgtct	trgaaagaac	agatotacat	attatatacc	craacrocra	5557555157
78680	aaaatgtcgg			tattaagaaa	cradarradc	
78620	ttgtcaatca	attacatgct	tetttgteae	tacattttt	ataacaaata	1111111111
78260	೭ ಡಿ೦೭ಡಿ೭ಡಿ ೪ ಡಿಡಿ	ttacttagat	במבבבבבב	gtttgtttgt	ttetatgttt	cadattcat
28200	ραδορρορρ	taaacactgg	ggccatttct	atgccaccgt	ttacatgaaa	taacaaqqct
28440	aggatttact	ddaaacctgt	ccctcaccca	aaggatgtga	gcctacttta	aagctaattt
28380	tddactgtat	atccaaaagt	rtdrdaagca	ataccccaaa	atacattgga	ttttaaaaaa
28320	rdagcaatga	ttttaaatat	ctacaatatt	gtttagaaag	tdcattctga	cgattaaccc
28260	эээдддггэд	ccsgcsgata	catccattt	crcagggatt	catttcacaa	tgatcactag
28200	creargrerg	· _	ಶತರತತತ್ತರಲ್ಲ	aaagtaagaa	rarradaara	tcsgaaagga
28740	בבבשבמבבב			cttcataaga		rødddac rac
28080		ассассасса	בכבבשכבבתב	tagtgtaaga	tettteaca	tgtctatgat
28020	даааассаас	астаттесст	acttgacatc	מבבבשבבבבם	tectaaacat	gagaagaatc
09672	tttgaaagat	rdrcaaaarc	авасасассд	tggtattaga	ctatcaaatg	cacaagtagc
27960	reacteteet		credercerr	cdraagatct	cagtgttctg	ದಿತ್ತಾರತಿಗಳ
27840	gattotott			agaaccattg		~
		aaatatcatg		כבכבכבבב		ctctctctct
27780	101021010	22222222		tccatggatc		attggacaca
27720		tggaattcca				dcffcadffc
09972		cacatgccac				atatggctga
00972		tacaagttca				_
27540		tgttaaatac tecaaattca		_		ttttacttac
27480		ctgtgctgaa '				_
27420		atacacacac				-
27360		_				~ ~ .
27300		grtrtttcaa : ctrgtttct :				
27240		aaatagacca :			•	
27180						
27120		catttetgea :			tgtcctgtga g	
27060		gartrgccgr				
27000		catcatactg :		, 1006100606		gaaccatcag
0₱692		rrccatggag (1 1161101016		ddcrccrdsd
26880	catataacaa	במבמכמבבבם ו	. erreeeree	. eeseesses	2 4422262322	raaracrarr c
26820	reactecet	actcadadda c	ימכדמכככדם :	t untopessor	sectdffddd G	רבשתשפת פרש
09762	ratectecaq	taraattta 1	ισοσεαασες ι	s estitetism	bigosogojo Joeofoolo	cctctctgtc o
00792	cacattcaat	rettiqqeae i	actecattt a	acercacee		
76640				scatgaacag s	atcggaatag a	
76580		ttataaatgc s		cacctgage o		
76520		ragtagoco a		gretecttt g		
09 ₹ 98	cagaatagac	cgtaaagtgt c		gattctaca g		
S6400				s epgatates	aattttett t	
0 7E9Z		ytatatttt s		caaaaaaga t		
26280		gegttaagat g			tgaaagtca s	
26220				ttgaagcta t	t teacttice	
26160	cagactctt	scrfddcagc f		ggttatctt c		_
26100				acttctgt a		
26040	sattaatgtc	grtttcatg a	:בבבבבבשש ב	:סבבבבבב ב	gagreratge t	
08652		: @@@@cc@cc	rrcaddrac r	catgtggct t		
25920	_	rraragece c	saagtagag t	laacactgga t	ettgatgatt g	tgttttaagc c

75..1 <255>

```
<223> potential microsequencing oligo 4-26-60.misl
                                                           £2..1 <252>
                                                      <221> primer_bind
                                          <223> pase G ; A in SEQ ID192
                                                               <222>
                                                         <221> allele
       <2223> polymorphic fragment 4-26-60, variant version of SEQ ID192
                                                           7£..£ <222>
                                                           <221> allele
                                                                 <550>
                                                     <213> Homo Sapiens
                                                              <SIS> DNY
                                                               L$ <TTZ>
                                                              <210> 269
                    attgtgcaga agttgccttt catgttcaaa aatgttaatt tgtttgt
L₹
                                                              897 <007>
         <223> complement potential microsequencing oligo 4-22-176.mis2
                                                           7£..22 <222>
                                                      <221> primer_bind
                     <223> potential microsequencing oligo 4-22-176.misi
                                                            5222> 1..23
                                                      <221> primer_bind
                                           <223> base G ; A in SEQ ID191
                                                               <222>
                                                           <221> allele
       <2223> polymorphic fragment 4-22-176, variant version of SEQ ID191
                                                            7£..1 <222>
                                                           <221> allele
                                                                  <5750>
                                                      <773> Howo 295; eua
                                                              <SIS> DNY
                                                               7717> 47
                                                               892 <012>
                     gaattgtgca gaagttgcct ttcctgttca aaaatgttaa tttgttt
LĐ
                                                               <$00 SQL
          <2223> complement potential microsequencing oligo 4-22-174 mis2
                                                            7£..25 <252>
                                                       <221> primer_bind
                     <223> potential microsequencing oligo 4-22-174.misl
                                                             £2..1 <222>
                                                       <221> primer_bind
                                           <223> base C ; A in SEQ ID190
                                                                <2225>
                                                            <221> allele
       <223> polymorphic fragment 4-22-174, variant version of SEQ ID190
                                                             ₹222> T..47
                                                            <221> allele
                                                      <213> Homo Sapiens
                                                               <212> DNA
                                                                <5115> 47
                                                               <510> 567
                     L₽
                                                               997 <000>
           <223> complement potential microsequencing oligo 4-20-77.mis2
                                                            7£..25 <222>
                                                       <221> primer_bind
                      <223> potential microsequencing oligo 4-20-77.misl
                                                       <222> primer_bind <222> 2..23
                                           <223> base T ; A in SEQ ID189
```

```
ZLZ <000>
          <223> complement potential microsequencing oligo 4-38-63.mis2
                                                             74..22 <222>
                                                       <221> primer_bind
                      <223> potential microsequencing oligo 4-38-63.misl
                                                              52..1 <222>
                                                        <221> primer_bind
                                           <223> pase G ; A in SEQ ID195
                                                                 <222>
                                                             <221> allele
       <2223> polymorphic fragment 4-38-63, variant version of SEQ ID195
                                                             <222> 4]1ele <222> 4]1ele <222>
                                                                    <220>
                                                       <213> Homo Sapiens
                                                                <SIS> DNY
                                                                 LD <TTZ>
                                                                 ZLZ <0TZ>
                     tattgggcct aaaacagtat tctgtaaagc ttaaattggt attaact
L₹
                                                                 TLZ <000>
           <223> complement potential microsequencing oligo 4-3-130.mis2
                                                              74..22 <222>
                                                        <221> primer_bind
                      <223> potential microsequencing oligo 4-3-130.misl
                                                               522.2>
                                                        <221> primer_bind
                                            <223> base G ; A in SEQ ID194
                                                                  77 <777>
                                                              <221> allele
        <2223> polymorphic fragment 4-3-130, variant version of SEQ ID194
                                                               74..1 <222>
                                                              <221> allele
                                                                     <550>
                                                        <213> Howo Sapiens
                                                                 <212> DNA
                                                                  LÞ <TTZ>
                                                                 <210> 211
                      catottaaga cagttagoag googaggago gactttaaag ggtgago
LĐ
                                                                 <400>
           <223> complement potential microsequencing oligo 4-26-72.mis2
                                                              7£..22 <222>
                                                         <221> primer_bind
                       <223> potential microsequencing oligo 4-26-72.misl
                                                               £2..1 <222>
                                                         <221> primer_bind
                                             <223> pase G # A in SEQ ID193
                                                                  DZ <ZZZ>
                                                              <221> allele
        <223> polymorphic fragment 4-26-72, variant version of SEQ ID193
                                                               7£..1 <222>
                                                              <221> allele
                                                                      <550>
                                                        <213> Homo Sapiens
                                                                  <515> DNY
                                                                  LÞ <TTZ>
                                                                  <210> 270
                      дагаддаваяд гусатства дасудтвус аууссааууа усуастт
 L₽
                                                                  < 4005 >
            <2223> complement potential microsequencing oligo 4-26-60.mis2
                                                               74... 25 <222>
                                                         <2221> primer_bind
```

```
<SIS> DNY
                                                                 L$ <112>
                                                                 912 <012>
                     сасада аасасудасу ауставава аааааааа ааааааа
L₽
                                                                 <400b>
           <223> complement potential microsequencing oligo 4-4-187.mis2
                                                             75... 25... 47
                                                        <221> primer_bind
                      <223> potential microsequencing oligo 4-4-187.misl
                                                               <222> 1..23
                                                        <221> primer_bind
                                            <223> base T ; A in SEQ ID198
                                                                  5777> 74
                                                              <221> allele
        <223> polymorphic fragment 4-4-187, variant version of SEQ ID198
                                                               7£..1 <222>
                                                              <221> allele
                                                                     <220>
                                                        <213> Homo Sapiens
                                                                 <212> DNA
                                                                  <5117> 47
                                                                 <210> 275
                      tactttccca ttpttcctga ctttgttatc ctatatataa acagaaa
L₽
                                                                 ラレて <00サ>
            <223> complement potential microsequencing oligo 4-4-152.mis2
                                                         <222> 25..47 | <
                       <223> potential microsequencing oligo 4-4-152.misl
                                                               5222>
                                                         <221> primer_bind <221>
                                             <223> base T ; C in SEQ ID197
                                                                  77 <772>
                                                              <221> allele
         <223> polymorphic fragment 4-4-152, variant version of SEQ ID197
                                                               74..1 <222>
                                                              <221> allele
                                                        <213> Homo Sapiens
                                                                 <ZIS> DNY
                                                                  <۲I۲> ط
                                                                  <210> 214
                      adcadagget asacttttt tttttgge aatgetgttg agaatat
L٥
                                                                  <400>
            <223> complement potential microsequencing oligo 4-38-83.mis2
                                                         <221> primer_bind <222> <222>
                       <223> potential microsequencing oligo 4-38-83.misl
                                                               £2..1 <222>
                                                         <221> primer_bind
                                             <223> base T | G in SEQ ID196
                                                                   4777> 74
                                                             <221> allele '
         <2223> polymorphic fragment 4-38-83, variant version of SEQ ID196
                                                               <221> allele
<222> 1..47
                                                                      <5250>
                                                         <213> Homo Sapiens
                                                                  <SIS> DNY
                                                                   LD <III>
                                                                  <210> 213
                      tataagttat aagaaaatca ggcggaggct aaacttttt tttgttt
 LĐ
```

```
7£... <222>
                                                              <221> allele
                                                                     <2220>
                                                       <213> Howo Sapiens
                                                                 <SIS> DNY
                                                                  <210> 270
                     гаадааадаа ттстдтдтс тддссааадт гтааасссас ададсса
L₹
                                                                 <400> 278
          <223> complement potential microsequencing oligo 4-42-401.mis2
                                                              72.. 25 .. 47
                                                         <2221> primer_bind
                      <223> potential microsequencing oligo 4-42-401.misl
                                                                52..1 <222>
                                                         <221> primer_bind
                                             <223> pase C ; A in SEQ ID201
                                                                   5777> 5
                                                              <221> allele
       <2223> polymorphic fragment 4-42-401, variant version of SEQ ID201
                                                               <222> allele <222> <222>
                                                                      <5250>
                                                         <213> Homo Sapiens
                                                                  <SIS> DNY
                                                                   45 <TTZ>
                                                                  8LZ <0TZ>
                      attatttaaa actatttatg taatcttatt ttcagggtt tttaatt
LÐ
                                                                  LLZ <000>
           <2223> complement potential microsequencing oligo 4-42-304 mis2
                                                              <222> 25..47 |
                                                          <SSI> primer_bind
                      <223> potential microsequencing oligo 4-42-304 misl
                                                                £2..1 <222>
                                                          <551> primer_bind
                                             <223> base T % C in SEQ ID200
                                                                   <222>
                                                               <221> allele
        <2223> polymorphic fragment 4-42-304, variant version of SEQ ID200
                                                                7£..1 <222>
                                                               <221> allele
                                                                      <2220>
                                                         <213> Homo Sapiens
                                                                   ANG <212>
                                                                    LD <TTZ>
                                                                   <210> 277
                       ctgtcatcaa ctaattttca caactaccta tgttttgatt tcatgta
 LĐ
                                                                   912 <000>
            <223> complement potential microsequencing oligo 4-4-288.mis2
                                                                7£..22 <222>
                                                          <221> primer_bind
                        <223> potential microsequencing oligo 4-4-288.misl
                                                          <221> primer_bind <222> 1..23
                                              <223> pase C | G in SEQ ID199
                                                                    ₹555> 5₹
                                                                <221> allele
          <2223> polymorphic fragment 4-4-288, variant version of SEQ ID199
                                                                 75..1 <255>
                                                                <221> allele
                                                                       <5250>
                                                         <213> Homo Sapiens
```

```
<223> polymorphic fragment 4-43-328, variant version of SEQ ID202
<221> allele
<222> 24
<223> base T ; C in SEQ ID202
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-43-328.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-43-328.mis2
<400> 279
                                                                       47
agaattctgt gttctggcca aagtttaaac ccacagagcc agtttaa
<210> 280
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-43-70, variant version of SEQ ID203
<221> allele
<222> 24
<223> base C ; G in SEQ ID203
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-43-70.mis1
<221> primer_bind
<222> 25..47 !
<223> complement potential microsequencing oligo 4-43-70.mis2
<400> 280
                                                                       47
atcgcctcca ttattctcaa aaacaccatg ggacacaaca caagaag
<210> 281
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-209, variant version of SEQ ID204
<221> allele
<222> 24
<223> base T ; C in SEQ ID204
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-209.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-209.mis2
<400> 281
                                                                        47
atatagagtg tgcatccctg acattgaaac tgaaggcttt atggttt
<210> 282
<211> 47
<212> DNA
<213> Homo Sapiens
 <220>
<221> allele
 <222> 1..47
 <223> polymorphic fragment 4-50-293, variant version of SEQ ID205
 <221> allele
 <222> 24
 <223> base T ; G in SEQ ID205
```

```
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-293.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-293.mis2
<400> 282
                                                                       47
cctgagtccc agggggctga cagtggacag tttaaaacat tgatgaa
<210> 283
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-323, variant version of SEQ ID206
<221> allele
<222> 24
<223> base T ; C in SEQ ID206
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-323.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-323.mis2
<400> 283
                                                                        47
tttaaaacat tgatgaatct ttattactac aaaagggttc gatttag
<210> 284
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-329, variant version of SEQ ID207
<221> allele
<222> 24
<223> base T ; C in SEQ ID207
<221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-50-329.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-50-329.mis2
 <400> 284
                                                                        47
 acattgatga attttatta ctataaaagg gttcgattta ggctagc
 <210> 285
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-50-33,0, variant version of SEQ ID208
 <221> allele
 <222> 24
 <223> base T ; A in SEQ ID208
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-50-330.mis1
 <221> primer_bind
```

```
<222> 25..47
<223> complement potential microsequencing oligo 4-50-330.mis2
                                                                       47
cattgatgaa totttattac tactaaaggg ttcgatttag gctagcc
<210> 286
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-52-163, variant version of SEQ ID209
<221> allele
<222> 24
<223> base C ; A in SEQ ID209
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-52-163.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-52-163.mis2
<400> 286
                                                                       47
gaacaggata ttcttaacta ccacagaatt ttacacatct attgttt
<210> 287
<211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-52-88, variant version of SEQ ID210
 <221> allele
 <222> 24
 <223> base T ; C in SEQ ID210
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-52-88.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-52-88.mis2
 <400> 287
                                                                        47
 tccatgtcat tattattcaa aagtttaaaa aatacacaag gtgaaaa
 <210> 288
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-53-258, variant version of SEQ ID211
 <221> allele
 <222> 24
 <223> base G ; A in SEQ ID211
  <221> primer_bind
 <222> 1..23
  <223> potential microsequencing oligo 4-53-258.mis1
  <221> primer_bind
  <222> 25..47
  <223> complement potential microsequencing oligo 4-53-258.mis2
  <400> 288
                                                                         47
  gagaaatcat gcagagagaa tgcgttctca ctcaaatttt aacctaa
```

```
<210> 289
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-54-283, variant version of SEQ ID212
<221> allele
<222> 24
<223> base T; A in SEQ ID212
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-54-283.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-54-283.mis2
                                                                       47
aagtagtttt tcacactttc tctttgatac aatcgatggc ttaatct
<210> 290
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-54-388, variant version of SEQ ID213
<221> allele
<222> 24
<223> base C ; A in SEQ ID213
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-54-388.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-54-388.mis2
                                                                        47
ctctctatcg tatacatctt tacccacgct gcagcgccaa gactcca
<210> 291
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-55-70, variant version of SEQ ID214
 <221> allele
 <222> 24
 <223> base T ; A in SEQ ID214
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-55-70.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-55-70.mis2
 <400> 291
                                                                         47
 tattaagaac ctaggtttta aaatactctc tatcgtatac atcttta
 <210> 292
 <211> 47
 <212> DNA
 <213> Homo Sapiens
```

```
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-55-95, variant version of SEQ ID215
<221> allele
<222> 24
<223> base C ; A in SEQ ID215
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-55-95.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-55-95.mis2
<400> 292
                                                                       47
ctctctatcg tatacatctt tacccacgct gcagcgccaa gactcca
<210> 293
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-56-159, variant version of SEQ ID216
<221> allele
<222> 24
<223> base T ; C in SEQ ID216
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-56-159.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-56-159.mis2
<400> 293
                                                                        47
aagttttcct tctcttctgt agatgtctcc atgttacagt caactat
<210> 294
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-56-213, variant version of SEQ ID217
<221> allele
 <222> 24
 <223> base G ; A in SEQ ID217
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-56-213.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-56-213.mis2
 <400> 294
                                                                        47
 atggctcatg ttcactctgg ttcgccttca gaggagtttg atatttt
 <210> 295
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> pclymorphic fragment 4-58-289, variant version of SEQ ID218
```

PCT/IB98/02133

```
<221> allele
<222> 24
<223> base C ; G in SEQ ID218
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-58-289.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-58-289.mis2
<400> 295
                                                                        47
catacctgca gcctgctttt ggtcaggggt gactacttta cctgcaa
<210> 296
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-58-318, variant version of SEQ ID219
<221> allele
<222> 24
<223> base C ; A in SEQ ID219
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-58-318.mis1
<221> primer_bind <222> 25..47
<223> complement potential microsequencing oligo 4-58-318.mis2
<400> 296
                                                                        47
tgactacttt acctgcaata tttctttgca agtttatttc ttccttt
<210> 297
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-60-266, variant version of SEQ ID220
<221> allele
<222> 24
<223> base T ; G in SEQ ID220
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-60-266.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-60-266.mis2
 <400> 297
                                                                         47
 aacaggacca agacactgca ttatataaag tttcagtatt tcttagc
 <210> 298
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-60-293, variant version of SEQ ID221
 <221> allele
 <222> 24
 <223> base T ; C in SEQ ID221
 <221> primer_bind
```

```
<222> 1..23
<223> potential microsequencing oligo 4-60-293.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-60-293.mis2
<400> 298
                                                                       47
aagtttcagt atttcttagc agatgaagcc agcaggaagt cctccta
<210> 299
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-84-241, variant version of SEQ ID222
<221> allele
<222> 24
<223> base T ; G in SEQ ID222
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-84-241.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-84-241.mis2
<400> 299
                                                                       47
gaaaaaaaaa tagtgactgc cactgtgaat aattcagttc ttcagaa
<210> 300
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-84-262, variant version of SEQ ID223
<221> allele
<222> 24
 <223> base G ; A in SEQ ID223
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-84-262.misl
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-84-262.mis2
 <400> 300
                                                                        47
 acggtgaata attcagttct tcagaagcag caacatgatc tcatgga
 <210> 301
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-86-206, variant version of SEQ ID224
 <221> allele
 <222> 24
 <223> base G ; A in SEQ ID224
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-86-206.mis1
 <221> primer_bind
 <222> 25..47 .
```

```
<223> complement potential microsequencing oligo 4-86-206.mis2
                                                                       47
gtattcaaat caggacacac cacgaatggc atctacacgt taacatt
<210> 302
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-86-309, variant version of SEQ ID225
<221> allele |
<222> 24
<223> base T ; A in SEQ ID225
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-86-309.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-86-309.mis2
                                                                        47
tggctctagg caggccactt tagtgagtga ggaaccagag agcagaa
<210> 303
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-88-349, variant version of SEQ ID226
<221> allele .
<222> 24
<223> base C ; G in SEQ ID226
<221> primer_bind
<222> 1..23
 <223> potential microsequencing oligo 4-88-349.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-88-349.mis2
 <400> 303
                                                                        47
 gaaactaaaa gacaatattc agtctgagat tttccaagtt ctttatg
 <210> 304
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-89-87, variant version of SEQ ID227
 <221> allele
 <222> 24
 <223> base T ; C in SEQ ID227
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-89-87.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-89-87.mis2
 <400> 304
                                                                         47
 ttcttccctg aacgctggtt tcatatagtt tttgtgttga gaataga
  <210> 305
```

```
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-123-184, variant version of SEQ ID228
<221> allele
<222> 24
<223> base C ; G in SEQ ID228
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-123-184.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-123-184.mis2
                                                                       47
ccagcccaga acattcacca gctcggccaa gagttctgct gggtttt
<210> 306
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-128-202, variant version of SEQ ID229
<221> allele
<222> 24
<223> base C ; A in SEQ ID229
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-128-202.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-128-202.mis2
                                                                        47
aatgtctgtt tcttagagaa ctgcaacaca cacacataca tacacac
<210> 307
<211> 47
<212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-128-275, variant version of SEQ ID230
 <221> allele
 <222> 24
 <223> base G ; A in SEQ ID230
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-128-275.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-128-275.mis2
 acacccctac ctcacatgtg taggcaaatg tatgcatata tgtctct
                                                                         47
 <210> 308
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
```

```
<221> allele
<222> 1..47
<223> polymorphic fragment 99-128-313, variant version of SEQ ID231
<221> allele
<222> 24
<223> base G ; A in SEQ ID231
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-128-313.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-128-313.mis2
<400> 308
                                                                        47
tatgtctcta gacagatata catgagattc tatttggcat agaaaaa
<210> 309
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-128-60, variant version of SEQ ID232
<221> allele
<222> 24
<223> base T ; C in SEQ ID232
<221> primer_bind
<222> 1..23
 <223> potential microsequencing oligo 99-128-60.mis1
<221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-128-60.mis2
· <400> 309
                                                                        47
 gcactgtgac ccaggcgcta ggttcctctt acagtgacac tccgaca
 <210> 310
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-12907-295, variant version of SEQ ID233
 <221> allele
 <222> 24
 <223> base G ; A in SEQ ID233
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-12907-295.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-12907-295.mis2
 <400> 310
                                                                        47
 gctatatggc attatatctc cacggggcag acctgatgta caagatg
 <210> 311
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-130-58, variant version of SEQ ID234
  <221> allele
```

```
<222> 24
<223> base T ; C in SEQ ID234
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-130-58.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-130-58.mis2
                                                                       47
aaagcaaaag agcttcaaaa atatttcagg agtgtgcata tggcgag
<210> 312
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-134-362, variant version of SEQ ID235
<221> allele .
<222> 24
<223> base T ; G in SEQ ID235
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-134-362.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-134-362.mis2
<400> 312
                                                                       47
caaaacactc atgttagtta gattattatt cctattacaa agataag
<210> 313
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-140-130, variant version of SEQ ID236
<221> allele
<222> 24
<223> base T ; C in SEQ ID236
<221> primer_bind
<222> 1..23
 <223> potential microsequencing oligo 99-140-130.mis1
<221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-140-130.mis2
 <400> 313
 tgttcaaaag cagctacaga ccatatgtaa acaattgagc atggctg
                                                                        47
 <210> 314
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-1462-238, variant version of SEQ ID237
 <221> allele
 <222> 24
 <223> base C ; G in SEQ ID237
 <221> primer_bind
 <222> 1..23
```

```
<223> potential microsequencing oligo 99-1462-238.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1462-238.mis2
<400> 314
                                                                        47
ccctttcaag gttagtaact catctgctgt gtttctgctt cagaagg
<210> 315
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-147-181, variant version of SEQ ID238
<221> allele
<222> 24
<223> base G ; A in SEQ ID238
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-147-181.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-147-181.mis2
<400> 315
                                                                        47
gtgtcatgaa aaagagcatg atagaaagaa aaacttaaat ctttata
<210> 316
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1474-156, variant version of SEQ ID239
<221> allele
<222> 24
<223> base T ; G in SEQ ID239
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1474-156.mis1
<221> primer_bind <222> 25..47
<223> complement potential microsequencing oligo 99-1474-156.mis2
                                                                        47
cttgtactca taagttaaat atttataaca agaagaaata tggactt
<210> 317
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-1474-359, variant version of SEQ ID240
 <221> allele
 <222> 24
 <223> base G ; A in SEQ ID240
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-1474-359.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-1474-359.mis2
```

```
<400> 317
                                                                       47
aaaaaaaatc aaattattgt accgaattcc ctaatatcag atgtgta
<210> 318
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1479-158, variant version of SEQ ID241
<221> allele
<222> 24
<223> base T ; C in SEQ ID241
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1479-158.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1479-158.mis2
<400> 318
                                                                        47
tttaaaaatc cacttgtaat cgctgctaat tggagtgtat attcagg
<210> 319
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1479-379, variant version of SEQ ID242
<221> allele
 <222> 24
 <223> base G ; A in SEQ ID242
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-1479-379.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-1479-379.mis2
 <400> 319
                                                                        47
 gtagagctgt gtactgaggt cagggaagca gctcatggta cagcctt
 <210> 320
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-148-129, variant version of SEQ ID243
 <221> allele
 <222> 24
 <223> base G ; A in SEQ ID243
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-148-129.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-148-129.mis2
 <400> 320
                                                                         47
 ttcatatcta tacaaataat tttgaattta atacataggg ctgcaaa
  <210> 321
  <211> 47
```

```
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-132, variant version of SEQ ID244
<221> allele
<222> 24
<223> base T ; C in SEQ ID244
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-132.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-143-132.mis2
<400> 321
                                                                       47
atatctatac aaataatttt gaatttaata catagggctg caaaaca
<210> 322
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-139, variant version of SEQ ID245
<221> allele
<222> 24
<223> base T ; C in SEQ ID245
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-139.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-139.mis2
                                                                        47
tacaaataat tttgaattta atatataggg ctgcaaaaca aggttga
<210> 323
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
 <222> 1..47
<223> polymorphic fragment 99-148-140, variant version of SEQ ID246
 <221> allele
 <222> 24
 <223> base G ; A in SEQ ID246
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-148-140.mis1
 <221> primer_bind
 <222> 25..47 |
 <223> complement potential microsequencing oligo 99-148-140.mis2
 <400> 323
                                                                         47
 acaaataatt ttgaatttaa tacgtagggc tgcaaaacaa ggttgat
 <210> 324
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
```

```
<222> 1..47
<223> polymorphic fragment 99-148-182, variant version of SEQ ID247
<221> allele
<222> 24
<223> base G ; A in SEQ ID247
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-182.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-182.mis2
<400> 324
                                                                       47
ttgatgttga tatgggcaac tgtgtgttgg atggtcccaa agcattc
<210> 325
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-366, variant version of SEQ ID248
<221> allele
<222> 24
<223> base T ; G in SEQ ID248
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-366.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-366.mis2
<400> 325
                                                                        47
teettgteaa aggtetetee etgttgetea eggetgeege eteaaag
<210> 326
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-148-76, variant version of SEQ ID249
 <221> allele
 <222> 24
 <223> base T ; C in SEQ ID249
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-148-76.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-148-76.mis2
 <400> 326
                                                                         47
 tgatagaatg ccttcctgaa ttattactct tgatggcttc ataaaac
 <210> 327
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
  <222> 1..47
  <223> polymorphic fragment 99-1480-290, variant version of SEQ ID250
  <221> allele
  <222> 24
```

```
<223> base T ; G in SEQ ID250
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1480-290.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1480-290.mis2
<400> 327
                                                                       47
tgcaccatct tcaccacaac ccctggcaac cactgatcct tttactg
<210> 328
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1481-285, variant version of SEQ ID251
<221> allele :
<222> 24
<223> base T ; G in SEQ ID251
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1481-285.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1481-285.mis2
<400> 328
                                                                       47
teccataace tgttttgett etetetetaa eeteaagatg gtataaa
<210> 329
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1484-101, variant version of SEQ ID252
<221> allele
<222> 24
<223> base C ; A in SEQ ID252
<221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-1484-101.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-1484-101.mis2
 <400> 329
                                                                        47
 aaaaagatca aatataagca tgtcactcct ctccttaaaa tctcagt
 <210> 330
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-1484-328, variant version of SEQ ID253
 <221> allele
 <222> 24
 <223> base C ; G in SEQ ID253
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-1484-328.mis1
```

```
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1484-328.mis2
                                                                       47
ggacacgtgg tcatgaggag tttcaaggga ttcagttttc agatccc
<210> 331
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1485-251, variant version of SEQ ID254
<221> allele
<222> 24
<223> base T ; G in SEQ ID254
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1485-251.misl
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1485-251.mis2
<400> 331
                                                                        47
gattgccttg atatatgctc ccatagaacc aagaatgtcc ccttttc
<210> 332
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1490-381, variant version of SEQ ID255
<221> allele
<222> 24
<223> base T |; C in SEQ ID255
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1490-381.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1490-381.mis2
<400> 332
                                                                        47
tgcacagtgg aaataccatg tcatggtacg ctactgtgca tctcttc
 <210> 333
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-1493-280, variant version of SEQ ID256
 <221> allele
 <222> 24
 <223> base G |; A in SEQ ID256
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-1493-280.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-1493-280.mis2
 <400> 333
```

```
47
ggatgacaga gtattgttgg agggatgggg tttggctgct tgttttt
<210> 334
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-151-94, variant version of SEQ ID257
<221> allele
<222> 24
<223> base G ; A in SEQ ID257
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-151-94.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-151-94.mis2
                                                                       47
attgagatca ttgataagga aatgttctaa aatttcaaaa tctatat
<210> 335
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-211-291, variant version of SEQ ID258
<221> allele
<222> 24
<223> base G ; A in SEQ ID258
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-211-291.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-211-291.mis2
<400> 335
                                                                        47
ctggttatat cagactgacc ttcgtgtttt caacaggtca atgcctt
<210> 336
<211> 46
<212> DNA
<213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..46
 <223> polymorphic fragment 99-213-37, variant version of SEQ ID259
 <221> allele
 <222> 23
 <223> base GC; T in SEQ ID259
 <221> primer_bind
 <222> 1..22
 <223> potential microsequencing oligo 99-213-37.mis1
 <221> primer_bind
 <222> 24..46
 <223> complement potential microsequencing oligo 99-213-37.mis2
 <400> 336
                                                                        46
 gtgcttccgg ctgcaggact gtgcggagga ctccagtgtc tgacag
 <210> 337
 <211> 47
 <212> DNA
```

```
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-221-442, variant version of SEQ ID260
<221> allele
<222> 24
<223> base C ; A in SEQ ID260
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-221-442.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-221-442.mis2
<400> 337
tgcctttgta gatatgcatg ggacttccat gacctagcca gacgaat
                                                                        47
<210> 338
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-222-109, variant version of SEQ ID261
<221> allele
<222> 24
<223> base T ; C in SEQ ID261
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-222-109.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-222-109.mis2
<400> 338
caggtgagga gtgctggatt ggctacgata tgaatttctt cagcagt
                                                                        47
<210> 339
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 185, SEQ 262, SEQ 186, SEQ 263,
SEQ 187, SEQ 264
<400> 339
tctaacctct catccaac
                                                                        18
<210> 340
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 188, SEQ 265, SEQ 189, SEQ 266
<400> 340
gttatcgtga gactttttc
                                                                        19
<210> 341
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
```

```
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 190, SEQ 267, SEQ 191, SEQ 268
<400> 341
                                                                        18
tgctggtgct gtgataac
<210> 342
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 192, SEQ 269, SEQ 193, SEQ 270
<400> 342
                                                                        18
tacagccctg taagacac
<210> 343
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 194, SEQ 271
<400> 343
                                                                         19
cagtatgttc aatgcacag
<210> 344
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 195, SEQ 272, SEQ 196, SEQ 273
<400> 344
                                                                         18
aaaacatcga catgggac
<210> 345
<211> 18
<212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> upstream amplification primer for SEQ 197, SEQ 274, SEQ 198, SEQ 275,
 SEQ 199, SEQ 276
 <400> 345
                                                                         18
 agcatttcga gtcatgtg
 <210> 346
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> upstream amplification primer for SEQ 200, SEQ 277, SEQ 201, SEQ 278
 <400> 346
                                                                          18
 ccctctttcc tcatgtag
 <210> 347
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
```

```
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 202, SEQ 279, SEQ 203, SEQ 280
<400> 347
                                                                       19
taactcgtaa acagagaac
<210> 348
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 204, SEQ 281, SEQ 205, SEQ 282,
SEQ 206, SEQ 283, SEQ 207, SEQ 284, SEQ 208, SEQ 285
<400> 348
                                                                        18
gcgtattgaa gctctttg
<210> 349
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 209, SEQ 286, SEQ 210, SEQ 287
<400> 349
                                                                        18
aacacgggga ttttaggc
<210> 350
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 211, SEQ 288
<400> 350
                                                                        19
cacatactaa ggctaatgg
<210> 351
<211> 18
<212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> upstream amplification primer for SEQ 212, SEQ 289, SEQ 213, SEQ 290
 <400> 351
                                                                        18
 gttgctggaa cctatttg
 <210> 352
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> upstream amplification primer for SEQ 214, SEQ 291, SEQ 215, SEQ 292
 <400> 352
                                                                         18
 tcgatggctt aatctacc
 <210> 353
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
```

WO 99/32644 PCT/IB98/02133

```
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 216, SEQ 293, SEQ 217, SEQ 294
<400> 353
                                                                        18
aaagaggagt aaatgggg
<210> 354
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 218, SEQ 295, SEQ 219, SEQ 296
<400> 354
                                                                        18
tccccacagc taagagcc
<210> 355
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 220, SEQ 297, SEQ 221, SEQ 298
<400> 355
                                                                         18
atacctaatt tcaggggg
<210> 356
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 222, SEQ 299, SEQ 223, SEQ 300
<400> 356
                                                                         19
ttaacagagt accttggag
<210> 357
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
 <221> primer_bind
 <222> 1..18
<223> upstream amplification primer for SEQ 224, SEQ 301, SEQ 225, SEQ 302
 <400> 357
                                                                         18
 gtacagcctt ttgcttac
 <210> 358
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> upstream amplification primer for SEQ 226, SEQ 303
 <400> 358
                                                                          18
 aacgtgtcat agaaagcc
 <210> 359
<211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
```

```
<222> 1..19
<223> upstream amplification primer for SEQ 227, SEQ 304
<400> 359
                                                                       19
gctgatgagt tagataacc
<210> 360
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 228, SEQ 305
<400> 360
                                                                        18
aaagccagga ctagaagg
<210> 361
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 229, SEQ 306, SEQ 230, SEQ 307,
SEQ 231, SEQ 308, SEQ 232, SEQ 309
<400> 361
                                                                        18
gaccagggtt taagttag
<210> 362
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 233, SEQ 310
<400> 362
                                                                        18
tctgttagga cctgtgag
<210> 363
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
 <222> 1..19
 <223> upstream amplification primer for SEQ 234, SEQ 311
 <400> 363
                                                                         19
 ccataacagc tagtacaac
 <210> 364
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> upstream amplification primer for SEQ 235, SEQ 312
 <400> 364
                                                                         18
 tggaaaggta ctcagaag
 <210> 365
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
```

```
<222> 1..19
<223> upstream amplification primer for SEQ 236, SEQ 313
<400> 365
                                                                        19
agagcatagt ataaagcag
<210> 366
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 237, SEQ 314
<400> 366
                                                                        19
ctagaagtag ctttaacag
<210> 367
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 238, SEQ 315
<400> 367
                                                                        19
gcagccaatc ttatatttc
<210> 368
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
 <222> 1..19
 <223> upstream amplification primer for SEQ 239, SEQ 316, SEQ 240, SEQ 317
 <400> 368
                                                                         19
 aaggttgtag agtagaaag
 <210> 369
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> upstream amplification primer for SEQ 241, SEQ 318, SEQ 242, SEQ 319
 <400> 369
                                                                         18
 caactgacac tataaccc
 <210> 370
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> upstream amplification primer for SEQ 243, SEQ 320, SEQ 244, SEQ 321,
 SEQ 245, SEQ 322, SEQ 246, SEQ 323, SEQ 247, SEQ 324, SEQ 248, SEQ 325, SEQ
 249, SEQ 326
  <400> 370
                                                                          18
  cagtggagtg tttatgtg
  <210> 371
  <211> 19
  <212> DNA
  <213> Homo Sapiens
  <220>
```

```
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 250, SEQ 327
<400> 371
                                                                       19
ttgcacaaaa ggtatagag
<210> 372
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 251, SEQ 328
<400> 372
                                                                        19
aggctcccct tttgagttg
<210> 373
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 252, SEQ 329, SEQ 253, SEQ 330
<400> 373
                                                                        18
atcctttcta gctgggag
<210> 374
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> upstream amplification primer for SEQ 254, SEQ 331
<400> 374
                                                                        20
 gtttaagaat gtgtgatggg
 <210> 375
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> upstream amplification primer for SEQ 255, SEQ 332
 <400> 375
                                                                        19
 aaggcaacag cgttgtgac
 <210> 376
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> upstream amplification primer for SEQ 256, SEQ 333
 <400> 376
                                                                         18
 ttttgggggt tttcagtg
 <210> 377
 <211> 18
 <212> DNA
  <213> Homo Sapiens
  <220>
  <221> primer_bind
```

```
<222> 1..18
<223> upstream amplification primer for SEQ 257, SEQ 334
<400> 377
                                                                        18
aacacaacag caaatccc
<210> 378
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> upstream amplification primer for SEQ 258, SEQ 335
<400> 378
                                                                        18
tccttacttg taaccccc
<210> 379
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> upstream amplification primer for SEQ 259, SEQ 336
<400> 379
                                                                         20
atactggcag cgtgtgcttc
<210> 380
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> upstream amplification primer for SEQ 260, SEQ 337
<400> 380
                                                                         19
ccctttttct tcactgttc
<210> 381
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
 <223> upstream amplification primer for SEQ 261, SEQ 338
 <400> 381
                                                                         20
 aggggagatg agggaagttg
 <210> 382
 <211> 20
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..20
 <223> downstream amplification primer for SEQ 185, SEQ 262, SEQ 186, SEQ 263,
 SEQ 187, SEQ 264
 <400> 382
                                                                         20
 gactgtatcc tttgatgcac
 <210> 383
 <211> 20
<212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
```

```
<222> 1..20
<223> downstream amplification primer for SEQ 188, SEQ 265, SEQ 189, SEQ 266
                                                                       20
gcataattgt gcttgactgg
<210> 384
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 190, SEQ 267, SEQ 191, SEQ 268
<400> 384
                                                                       18
tgctgagagg agcttttg
<210> 385
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 192, SEQ 269, SEQ 193, SEQ 270
<400> 385
                                                                        18
tgaggactgc taggaaag
<210> 386
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 194, SEQ 271
<400> 386
                                                                        20
acaaatcag gaacaatggg
 <210> 387
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> downstream amplification primer for SEQ 195, SEQ 272, SEQ 196, SEQ 273
 <400> 387
                                                                        18
 ttgcattttc cccccaac
 <210> 388
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> downstream amplification primer for SEQ 197, SEQ 274, SEQ 198, SEQ 275,
 SEQ 199, SEQ 276
 <400> 388
                                                                         18
 accatttgga caatgggg
 <210> 389
 <211> 20
 <212> DNA
 <213> Homo Sapiens
  <220>
  <221> primer_bind
```

```
<222> 1..20
<223> downstream amplification primer for SEQ 200, SEQ 277, SEQ 201, SEQ 278
<400> 389
                                                                        20
gctcttaaac tggctctgtg
<210> 390
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 202, SEQ 279, SEQ 203, SEQ 280
<400> 390
                                                                        18
ggcatgactt cacgtttc
<210> 391
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 204, SEQ 281, SEQ 205, SEQ 282,
SEQ 206, SEQ 283, SEQ 207, SEQ 284, SEQ 208, SEQ 285
<400> 391
                                                                        18
aggatettet acagteae
<210> 392
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 209, SEQ 286, SEQ 210, SEQ 287
<400> 392
                                                                        20
tggtagcgtt tgaaatcatc
<210> 393
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 211, SEQ 288
<400> 393
                                                                        20
tataagcaca aataggttcc
 <210> 394
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> downstream amplification primer for SEQ 212, SEQ 289, SEQ 213, SEQ 290
 <400> 394
                                                                         18
 gaataactga ggggagtg
 <210> 395
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
```

```
<222> 1..19
<223> downstream amplification primer for SEQ 214, SEQ 291, SEQ 215, SEQ 292
                                                                       19
gtgaatctcc ttttccaag
<210> 396
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 216, SEQ 293, SEQ 217, SEQ 294
<400> 396
                                                                       18
ctaaggtgtt gtagacag
<210> 397
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 218, SEQ 295, SEQ 219, SEQ 296
<400> 397
                                                                       20
cacctcgata aatcaagtcc
<210> 398
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 220, SEQ 297, SEQ 221, SEQ 298
<400> 398
                                                                        20
gttcacttaa ttctgttgag
<210> 399
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> downstream amplification primer for SEQ 222, SEQ 299, SEQ 223, SEQ 300
 <400> 399
                                                                        18
 cgccttttct gaaaggtg
 <210> 400
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> downstream amplification primer for SEQ 224, SEQ 301, SEQ 225, SEQ 302
 <400> 400
                                                                         18
 attttctgca cagcagcg
 <210> 401
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
  <222> 1..19
```

```
<223> downstream amplification primer for SEQ 226, SEQ 303
<400> 401
                                                                       19
tattttctag ctcttctgg
<210> 402
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> downstream amplification primer for SEQ 227, SEQ 304
<400> 402
                                                                       19
agcaagagtg attgtaaag
<210> 403
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 228. SEQ 305
<400> 403
                                                                       18
tattcagaaa ggagtggg
<210> 404
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 229, SEQ 306, SEQ 230, SEQ 307,
SEQ 231, SEQ 308, SEQ 232, SEQ 309
<400> 404
                                                                        18
agagcgttct tgcctttc
<210> 405
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 233, SEQ 310
<400> 405
                                                                        20
ggtaacccta aaatgttatc
 <210> 406
 <211> 21
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..21
 <223> downstream amplification primer for SEQ 234, SEQ 311
 <400> 406
                                                                        21
 agaaaccata agggtatatt g
 <210> 407
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
```

```
<223> downstream amplification primer for SEQ 235, SEQ 312
<400> 407
                                                                       19
acagtgcaaa ggttatatc
<210> 408
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 236, SEQ 313
<400> 408
                                                                        21
gaacaacctt gaattagctt g
<210> 409
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 237, SEQ 314
<400> 409
                                                                        21
gattccagaa gtccatttca g
<210> 410
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 238, SEQ 315
<400> 410
                                                                        21
aggtaagaat gagcaaaaag g
 <210> 411
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> downstream amplification primer for SEQ 239, SEQ 316, SEQ 240, SEQ 317
 <400> 411
                                                                         19
 gcttgtgttt gttcaattc
 <210> 412
 <211> 18
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..18
 <223> downstream amplification primer for SEQ 241, SEQ 318, SEQ 242, SEQ 319
 <400> 412
                                                                         18
 cttgaaatac tcccagcc
 <210> 413
 <211> 19
 <212> DNA
 <213> Homo Sapiens
  <220>
  <221> primer_bind
  <222> 1..19
```

```
<223> downstream amplification primer for SEQ 243, SEQ 320, SEQ 244, SEQ 321,
SEQ 245, SEQ 322, SEQ 246, SEQ 323, SEQ 247, SEQ 324, SEQ 248, SEQ 325, SEQ
249, SEQ 326
<400> 413
                                                                        19
ccatgaactg agaactttg
<210> 414
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 250, SEQ 327
                                                                        18
ggtgacaggt aaagaaac
<210> 415
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 251, SEQ 328
<400> 415
                                                                        21
attcaggcac agaagtcata c
<210> 416
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 252, SEQ 329, SEQ 253, SEQ 330
<400> 416
                                                                        21
agggcagcac aatgtagtaa g
<210> 417
<211> 18
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..18
<223> downstream amplification primer for SEQ 254, SEQ 331
 <400> 417
                                                                         18
 cctctttatc tccaaacc
 <210> 418
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> downstream amplification primer for SEQ 255, SEQ 332
 <400> 418
                                                                         19
 gaaaacaatc aagctctgg
 <210> 419
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
```

```
<222> 1..19
<223> downstream amplification primer for SEQ 256, SEQ 333
<400> 419
                                                                       19
cctttatatc cttggagtc
<210> 420
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 257, SEQ 334
<400> 420
                                                                       21
tattacacgt tccaactctt c
<210> 421
<211> 20
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..20
<223> downstream amplification primer for SEQ 258, SEQ 335
<400> 421
                                                                        20
ctgtgtttaa gtgactgctg
<210> 422
<211> 21
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..21
<223> downstream amplification primer for SEQ 259, SEQ 336
<400> 422
                                                                        21
 ttattgcccc acatgcttga g
 <210> 423
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> downstream amplification primer for SEQ 260, SEQ 337
 <400> 423
                                                                        19
 tcattcgtct ggctaggtc
 <210> 424
 <211> 21
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..21
 <223> downstream amplification primer for SEQ 261, SEQ 338
 <400> 424
                                                                         21
 gaaacagact gaagcaagga c
 <210> 425
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
```

WO 99/32644 PCT/IB98/02133

```
<223> potential microsequencing oligo for 4-14-107.mis1
<400> 425
                                                                        19
acaaccacca aatgcatac
<210> 426
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-14-317.mis1
<400> 426
                                                                        19
acatgcaagg tgggcaaga
<210> 427
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-14-35.misl
                                                                        19
aacacagaaa ccgctaaaa
<210> 428
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-20-149.mis1
<400> 428
                                                                        23
tttttgctgt gtcttcaaag tga
 <210> 429
<211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 4-20-77.mis1
 <400> 429
                                                                         19
 acatgaagat tctgaaggg
 <210> 430
 <211> 23
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..23
 <223> microsequencing oligo for 4-22-174.mis1
 <400> 430
                                                                         23
 ggattgtgca gaagttgcct ttc
 <210> 431
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 4-22-176.mis1
```

```
<400> 431
                                                                       19
tgcagaagtt gcctttcat
<210> 432
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-26-60.mis1
<400> 432
                                                                       19
ggaaagtgca tcttaagac
<210> 433
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-26-72.mis1
<400> 433
                                                                        19
ttaagacagt tagcaggcc
<210> 434
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-3-130.mis1
<400> 434
                                                                        19
gggcctaaaa cagtattct
 <210> 435
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 4-38-63.mis1
 <400> 435
                                                                        19
 agttataaga aaatcaggc
 <210> 436
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 4-38-83.mis1
 <400> 436
                                                                         19
 gaggctaaac tttttttt
  <210> 437
  <211> 19
  <212> DNA
  <213> Homo Sapiens
  <220>
  <221> primer_bind
  <222> 1..19
  <223> potential microsequencing oligo for 4-4-152.mis1
  <400> 437
```

}	
ttcccattgt tcctgactt	19
<210> 438	
<211> 23	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<pre><222> 123 <223> microsequencing oligo for 4-4-187.mis1</pre>	
<400> 438	
tataaacaga aacatggatg agt	23
<210> 439	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> potential microsequencing oligo for 4-4-288.misl	
<400> 439	19
catcaactaa ttttcacaa	
<210> 440 <211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> potential microsequencing oligo for 4-42-304.mis1	
<400> 440	19
tttaaaacta tttatgtaa	19
<210> 441	
<211> 23	
<212> DNA	
<213> Homo Sapiens <220>	
<221> primer_bind	
<222> 123	
<223> microsequencing oligo for 4-42-401.mis1	
<400> 441	
taagaaagaa ttctgtgttc tgg	23
<210> 442	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<pre><221> primer_bind</pre>	
<pre><222> 119 <223> potential microsequencing oligo for 4-43-328.mis1</pre>	
<400> 442	
ttctgtgttc tggccaaag	19
<210> 443	
<211> 23	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 123	
<223> microsequencing oligo for 4-43-70.mis1	
<400> 443	23
atcgcctcca ttattctcaa aaa	

```
<210> 444
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-50-209.mis1
<400> 444
                                                                        23
atatagagtg tgcatccctg aca
<210> 445
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-50-293.mis1
<400> 445
                                                                        23
cctgagtccc agggggctga cag
<210> 446
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-50-323.mis1
 <400> 446
                                                                        23
 tttaaaacat tgatgaatct tta
 <210> 447
 <211> 23
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..23
 <223> microsequencing oligo for 4-50-329.mis1
 <400> 447
                                                                         23
 acattgatga atctttatta cta
 <210> 448
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 4-50-330.mis1
 <400> 448
                                                                         19
  gatgaatctt tattactac
  <210> 449
  <211> 23
  <212> DNA
  <213> Homo Sapiens
  <220>
  <221> primer_bind
  <222> 1..23
  <223> microsequencing oligo for 4-52-163.mis1
  <400> 449
                                                                          23
  gaacaggata ttcttaacta cca
  <210> 450
```

```
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-52-88.mis1
<400> 450
                                                                       23
tccatgtcat tattattcaa aag
<210> 451
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-53-258.misl
<400> 451
                                                                       19
aatcatgcag agagaatgc
<210> 452
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-54-283.mis1
<400> 452
                                                                        23
aagtagtttt tcacactttc tct
<210> 453
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
 <222> 1..19
<223> potential microsequencing oligo for 4-54-388.mis1
 <400> 453
                                                                        19
 ctatcgtata catctttac
 <210> 454
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 4-55-70.mis1
 <400> 454
                                                                        19
 aagaacctag gttttaaaa
 <210> 455
 <211> 23
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..23
 <223> microsequencing oligo for 4-55-95.mis1
  <400> 455
                                                                         23
 ctctctatcg tatacatctt tac
  <210> 456
  <211> 23
```

<pre><212> DNA</pre>	
<pre><213> Homo Sapiens</pre>	
<pre><221> primer_bind</pre>	
<pre><222> 123 <223> microsequencing oligo for 4-56-159.mis1</pre>	
<pre><223> microsequencing origo for 4 30 133.mis1</pre>	
agttttcct tctcttctgt aga	23
<210> 457	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220> <221> primer_bind	
<222 > 1 19	
<223> potential microsequencing oligo for 4-56-213.mis1	
<400> 457	19
ctcatgttca ctctggttc	19
<210> 458	
<211> 23	
<212> DNA <213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 123	
<223> microsequencing oligo for 4-58-289.mis1	
<400> 458	23
catacctgca gcctgctttt ggt	
<210> 459 <211> 23	
<211> 23 <212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 123	
<223> microsequencing oligo for 4-58-318.mis1	
<400> 459 tgactacttt acctgcaata ttt	23
<210> 460	
<211> 23	
<212> DNA	
<213> Homo Sapiens	
<220>	
<pre><221> primer_bind</pre>	
<222> 123 <223> microsequencing oligo for 4-60-266.mis1	
<400> 460	
aacaggacca agacactgca tta	23
<210> 461	
<211> 23	
<212> DNA	
<213> Homo Sapiens <220>	
<221> primer_bind .	
<222> 123	
<223> microsequencing oligo for 4-60-293.mis1	
<400> 461	23
aagtttcagt atttcttagc aga	20
<210> 462 <211> 19	
<211> IS <212> DNA	

```
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-84-241.mis1
<400> 462
                                                                        19
aaaaaatagt gactgccac
<210> 463
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind <222> 1..19
<223> potential microsequencing oligo for 4-84-262.mis1
<400> 463
                                                                        19
tgaataattc agttcttca
<210> 464
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-86-206.misl
<400> 464
                                                                         19
tcaaatcagg acacaccac
<210> 465
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-86-309.mis1
<400> 465
                                                                         19
 tctaggcagg ccactttag
 <210> 466
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 4-88-349.mis1
 <400> 466
                                                                         19
 ctaaaagaca atattcagt
 <210> 467
 <211> 23
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..23
 <223> microsequencing oligo for 4.89-87.mis1
 <400> 467
                                                                          23
  ttcttccctg aacgctggtt tca
  <210> 468
  <211> 19
  <212> DNA
  <213> Homo Sapiens
```

```
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-123-184.mis1
                                                                        19
cccagaacat tcaccagct
<210> 469
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-128-202.mis1
<400> 469
                                                                        19
tctgtttctt agagaactg
<210> 470
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-128-275.mis1
<400> 470
                                                                        19
ccctacctca catgtgtag
<210> 471
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-128-313.mis1
<400> 471
                                                                        19
tctctagaca gatatacat
<210> 472
<211> 23
<212> DNA
<213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..23
 <223> microsequencing oligo for 99-128-60.mis1
 <400> 472
                                                                         23
 cactgtgacc caggcgctag cgt
 <210> 473
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing qligo for 99-12907-295.mis1
 <400> 473
                                                                         19
 tatggcatta tatctccac
 <210> 474
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
```

```
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-130-58.mis1
<400> 474
                                                                        19
caaaagagct tcaaaaata
<210> 475
<211> 19
<212> ENA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-134-362.mis1
<400> 475
                                                                        19
acactcatgt tagttagat
<210> 476
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-140-130.mis1
<400> 476
                                                                        19
caaaagcagc tacagacca
<210> 477
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1462-238.mis1
<400> 477
                                                                         19
ttcaaggtta gtaactcat
<210> 478
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 99-147-181.mis1
 <400> 478
                                                                         19
 catgaaaaag agcatgata
 <210> 479
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 99-1474-156.mis1
 <400> 479
                                                                         19
 tactcataag ttaaatatt
 <210> 480
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
```

```
<222> 1..19
<223> potential microsequencing oligo for 99-1474-359.mis1
<400> 480
                                                                        19
aaaatcaaat tattgtacc
<210> 481
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1479-158.mis1
<400> 481
                                                                        19
aaaatccact tgtaatcgc
<210> 482
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1479-379.mis1
<400> 482
                                                                        19
agctgtgtac tgaggtcag
<210> 483
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
 <222> 1..19
<223> potential microsequencing oligo for 99-148-129.mis1
 <400> 483
                                                                         19
 tatctataca aataatttt
 <210> 484
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 99-148-132.mis1
 <400> 484
                                                                         19
 ctatacaaat aattttgaa
 <210> 485
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 99-148-139.mis1
 <400> 485
                                                                         19
  aataattttg aatttaata
  <210> 486
  <211> 19
  <212> DNA
  <213> Homo Sapiens
  <220>
  <221> primer_bind
  <222> 1..19
```

WO 99/32644 "PCT/IB98/02133

<223> potential microsequencing oligo for 99-140-140.mls1	
<400> 486	4.0
ataattttga atttaatac	19
<210> 487	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> potential microsequencing oligo for 99-148-182.mis1	
<400> 487	
tgttgatatg ggcaactgt	19
<210> 488	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> potential microsequencing oligo for 99-148-366.mis1	
<400> 488	
tgtcaaaggt ctctccctg	19
<210> 489	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> primer_brid <222> 119	
<pre><222> 119 <223> potential microsequencing oligo for 99-148-76.mis1</pre>	
<400> 489	19
agaatgcctt cctgaatta	
<210> 490 <211> 19	
<211> 19 <212> DNA	
<213> Homo Sapiens	
-	
<220>	
<pre><221> primer_bind <222> 119</pre>	
<pre><222> 119 <223> potential microsequencing oligo for 99-1480-290.mis1</pre>	
<400> 490	
	19
ccatcttcac cacaacccc	
<210> 491 <211> 19	
<211> 19 <212> DNA	
,	
<213> Homo Sapiens	
<220> <221> primer_bind	
•	
<222> 119 <223> potential microsequencing oligo for 99-1481-285.mis1	
<400> 491	19
ataacctgtt ttgcttctc	
<210> 492	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<pre><222> 119 <223> potential microsequencing oligo for 99-1484-101.mis1</pre>	
ZZZJY DOUGHILIAI MILOIOSEGGENCING OLIGO IVI JY IIV I IIII	

```
<400> 492
                                                                       19
agatcaaata taagcatgt
<210> 493
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1484-328.mis1
<400> 493
                                                                        19
acgtggtcat gaggagttt
<210> 494
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1485-251.mis1
<400> 494
                                                                        19
gccttgatat atgctccca
<210> 495
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1490-381.mis1
<400> 495
                                                                        19
cagtggaaat accatgtca
<210> 496
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bina
<222> 1..19
<223> potential microsequencing oligo for 99-1493-280.mis1
 <400> 496
                                                                         19
 gacagagtat tgttggagg
 <210> 497
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 99-151-94.mis1
 <400> 497
                                                                         19
 agatcattga taaggaaat
 <210> 498
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> microsequencing oligo for 99-211-291.mis1
 <400> 498
```

WO 99/32644 PCT/IB98/02133

ttatatcaga ctgaccttc <210> 499 <211> 19 <212> DNA	19
<213> Homo Sapiens <220>	
<pre><221> primer_bind <222> 119</pre>	
<223> potential microsequencing oligo for 99-213-37.mis1 <400> 499	19
cttccggctg caggactgt <210> 500	+7
<211> 19	
<212> DNA <213> Homo Sapiens	
<220>	
<221> primer_bind	
<pre><222> 119 <223> potential microsequencing oligo for 99-221-442.mis1</pre>	
<400> 500	4.0
tttgtagata tgcatggga	19
<210> 501 <211> 23	
<212> DNA	
<213> Homo Sapiens	
<220>	
<pre><221> primer_bind <222> 123</pre>	
<223> microsequencing oligo for 99-222-109.mis1	
<400> 501	2.2
caggtgagga gtgctggatt ggc	23
<210> 502 <211> 23	
<212> DNA	
<213> Homo Sapiens	
<220>	
<pre><221> primer_bind <222> 123</pre>	
<223> microsequencing oligo for 4-14-107.mis2	
<400> 502	23
ctatcaggca tttgcctggt tgc <210> 503	2.5
<211> 23	
<212> DNA	
<213> Homo Sapiens	
<220> <221> primer_bind	
<222> 123	
<223> microsequencing oligo for 4-14-317.mis2	
<400> 503 tcatgacctg tgcccacctc ttt	23
<210> 504	
<211> 23	
<212> DNA ,	
<213> Homo Sapiens (220>	
<221> primer_bind	
<222> 123 ;	
<223> microsequencing oligo for 4-14-35.mis2	
tctctgcaga cagcttctgc ctg	23
\cdot	

```
<210> 505
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-20-149.mis2
                                                                       19
agcaggcaat aaaccaaga
<210> 506
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-20-77.mis2
<400> 506
                                                                       19
gtgttctcag acaacaaag
<210> 507
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-22-174.mis2
<400> 507
                                                                        19
aaattaacat ttttgaaca
<210> 508
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-22-176.mis2
<400> 508
                                                                        19
acaaattaac atttttgaa
<210> 509
<211> 23
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..23
 <223> microsequencing oligo for 4-26-60.mis2
 <400> 509
                                                                        23
 aagtcgctcc tcggcctgct aac
 <210> 510
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 4-26-72.mis2
 <400> 510
                                                                         19
 accetttaaa gtegeteet
 <210> 511
```

<211> 23			
<212> DNA <213> Homo Sapi	ens		
<213> Homo Bap:	CIID		
<221> primer_bi	.nd		
<222> 123			
<223> microsequ	encing oligo for $4-3$	-130.mis2	
<400> 511		23	
agttaatacc aatt	taagct tta	23	
<210> 512			
<211> 19			
<212> DNA			
<213> Homo Sap: <220>	.ens		
<220> <221> primer_b:	ind		
<221> primer_b.			
<223> notentia	l microsequencing oli	lgo for 4-38-63.mis2	
<400> 512			
aaaaaaaaag ttt	ageete	19	
<210> 513			
<211> 23			
<212> DNA			
<213> Homo Sap	iens		
<220>			
<221> primer_b	ına		
<222> 123	uencing oligo for 4-	38-83 mis2	
<223> microseq <400> 513	lencing origo for 4-	50 03.MIDE	
atattctcaa cag	cattocc aaa	23	
<210> 514	categoo aaa		
<211> 19			
<212> DNA			
<213> Homo Sap	iens		
<220>			
<221> primer_b	ind		
<222> 119		for 4-4-152 mis2	
	1 microsequencing ol	1go 101 4-4-152.m152	
<400> 514		19)
tgtttatata tag	gataac		
<211> 515			
<211> 19 <212> DNA			
<213> Homo Sap	oiens		
<220>			
<221> primer_k	oind		
<222> 119	_		
<223> potentia	al microsequencing ol	igo for 4-4-187.mis2	
<400> 515		1	9
ttttttttt tt	ttttt	*	_
<210> 516			
<211> 19			
<212> DNA <213> Homo Say	niens		
<213> Hollio Saj	Jiens .		
<221> primer_	bind		
<222> 119	•		
<223> potenti	al microsequencing o	ligo for 4-4-288.mis2	
<400> 516	:		.9
tgaaatcaaa ac	ataggta	1	. 7
<210> 517			
<211> 19			

```
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-42-304.mis2
<400> 517
                                                                       19
aaaaacccct gaaaataag
<210> 518
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-42-401.mis2
<400> 518
                                                                        19
tctgtgggtt taaactttg
<210> 519
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-43-328.mis2
<400> 519
                                                                        19
actggctctg tgggtttaa
<210> 520
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-43-70.mis2
 <400> 520
                                                                        19
ttgtgttgtg tcccatggt
 <210> 521
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 4-50-209.mis2
 <400> 521
                                                                         19
 cataaagcct tcagtttca
 <210> 522
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 4-50-293.mis2
 <400> 522
                                                                         19
 tcaatgtttt aaactgtcc
 <210> 523
 <211> 19
 <212> DNA
```

```
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-50-323.mis2
<400> 523
                                                                         19
atcgaaccct tttgtagta
<210> 524
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-50-329.mis2
<400> 524
                                                                         19
gcctaaatcg aaccctttt
<210> 525
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-50-330.mis2
<400> 525
                                                                         19
agcctaaatc gaacccttt
<210> 526
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-52-163.mis2
<400> 526
                                                                          19
aatagatgtg taaaattct
<210> 527
<211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
<221> primer_bind <222> 1..19
 <223> potential microsequencing oligo for 4-52-88.mis2
 <400> 527
                                                                          19
 caccttgtgt atttttaa
 <210> 528
 <211> 23
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..23
 <223> microsequencing oligo for 4-53-258.mis2
 <400> 528
                                                                          23
 ttaggttaaa atttgagtga gaa
 <210> 529
<211> 19
 <212> DNA
 <213> Homo Sapiens
```

```
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-54-283.mis2
<400> 529
                                                                       19
taagccatcg attgtatca
<210> 530
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-54-388.mis2
<400> 530
                                                                        19
gtcttggcgc tgcagcgtg
<210> 531
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 4-55-70.mis2
<400> 531
                                                                        23
taaagatgta tacgatagag agt
<210> 532
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-55-95.mis2
<400> 532
                                                                        19
gtcttggcgc tgcagcgtg
<210> 533
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 4-56-159.mis2
 <400> 533
                                                                        19
 ttgactgtaa catggagac
 <210> 534
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 4-56-213.mis2
 <400> 534
                                                                         19
 tatcaaactc ctctgaagg
 <210> 535
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
```

```
<221> primer_bind
<222> 1..19 '
<223> potential microsequencing oligo for 4-58-289.mis2
<400> 535
                                                                        19
aggtaaagta gtcacccct
<210> 536
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-58-318.mis2
<400> 536
                                                                        19
gaagaaataa acttgcaaa
<210> 537
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-60-266.mis2
<400> 537
                                                                        19
agaaatactg aaactttat
<210> 538
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 4-60-293.mis2
<400> 538
                                                                        19
aggacttcct gctggcttc
<210> 539
<211> 23
<212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..23
 <223> microsequencing oligo for 4-84-241.mis2
 <400> 539
                                                                         23
 ttctgaagaa ctgaattatt cac
 <210> 540
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 4-84-262.mis2
 <400> 540
                                                                         19
 tgagatcatg ttgctgctt
 <210> 541
 <211> 23
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
```

<222> 123 <223> microsequencing oligo for 4-86-206.mis2	
<400> 541 : aatgttaacg tgtagatgcc att <210> 542	23
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> potential microsequencing oligo for 4-86-309.mis2	
<400> 542	19
gctctctggt tcctcactc	19
<210> 543	
<211> 19	
<212> DNA <213> Homo Sapiens	
<220>	
<220> <221> primer_bind	
<222> 119	
<223> potential microsequencing oligo for 4-88-349.mis2	
<400> 543	
aagaacttgg aaaatctca	19
<210> 544	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<pre><222> 119 <223> potential microsequencing oligo for 4-89-87.mis2</pre>	
<223> potential microsequencing origo for 4-05-07.miss	
ttctcaacac aaaaactat	19
<210> 545	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<pre><223> potential microsequencing oligo for 99-123-184.mis2</pre>	
<400> 545 cccagcagaa ctcttggcc	19
<210> 546	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> potential microsequencing oligo for 99-128-202.mis2	
<400> 546	19
gtatgtatgt gtgtgttt ,	
<210> 547 <211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	

<223> potential microsequencing origo for 99-128-273.mrsz	
<400> 547	19
acatatatgc atacatttg <210> 548	
<211> 19	
<211> 19 <212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> potential microsequencing oligo for 99-128-313.mis2	
<400> 548	
tctatgccaa atagaatct	19
<210> 549	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> potential microsequencing oligo for 99-128-60.mis2	
<400> 549	19
ggagtgtcac tgtaagagg	1)
<210> 550	
<211> 19 <212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> microsequencing oligo for 99-12907-295.mis2	
<400> 550	
ttgtacatca ggtctgccc	19
<210> 551	
<211> 23	
<212> DNA	
<213> Homo Sapiens	
<220>	
<pre><221> primer_bind <222> 123</pre>	
<pre><222> 123 <223> microsequencing oligo for 99-130-58.mis2</pre>	
<400> 551	
ctcgccatat gcacactcct gaa	23
<210> 552	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> microsequencing oligo for 99-134-362.mis2	
<400> 552	19
tctttgtaat aggaataat	13
<210> 553 <211> 19	
<211> 19	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> microseguencing oligo for 99-140-130.mis2	

, · .

```
<400> 553
                                                                        19
catgctcaat tgtttacat
<210> 554
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1462-238.mis2 <\!400\!> 554 :
                                                                         19
ctgaagcaga aacacagca
<210> 555
<211> 23
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..23
<223> microsequencing oligo for 99-147-181.mis2
<400> 555
                                                                         23
tataaagatt taagtttttc ttt
<210> 556
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1474-156.mis2
<400> 556
                                                                         19
ccatatttct tcttgttat
<210> 557
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1474-359.mis2
<400> 557
                                                                         19
catctgatat tagggaatt
<210> 558
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1479-158.mis2
 <400> 558
                                                                         19
aatatacact ccaattagc
 <210> 559
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 99-1479-379.mis2
 <400> 559
```

ctgtaccatg agctgcttc <210> 560 <211> 19	19
<211> 15 <212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> primer_brid <222> 119	
<222> 119 <223> potential microsequencing oligo for 99-148-129.mis2	
<400> 560	19
cagccctatg tattaaatt	+2
<210> 561	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> potential microsequencing oligo for 99-148-132.mis2	
<400> 561	19
ttgcagccct atgtattaa	17
<210> 562	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> potential microsequencing oligo for 99-148-139.mis2	
<400> 562	19
ccttgttttg cagccctat	+7
<210> 563	
<211> 19	
<212> DNA	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> potential microsequencing oligo for 99-148-140.mis2	
<400> 563	19
accttgtttt gcagcccta	+7
<210> 564	
<211> 23	
<212> DNA	
<213> Homo Sapiens	
<220> <221> primer_bind	
<222> primer_brind <222> 123	
<pre><222> 123 <223> microsequencing oligo for 99-148-182.mis2</pre>	
<400> 564	
gaatgetttg ggaccateca aca	23
<210> 565	
<211> 19	
<212> DNA .	
<213> Homo Sapiens	
<220>	
<221> primer_bind	
<222> 119	
<223> potential microsequencing oligo for 99-148-366.mis2	
<400> 565	
gaggcggcag ccgtgagca	19

. .

```
<210> 566
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-148-76.mis2
<400> 566
                                                                        19
tatgaagcca tcaagagta
<210> 567
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1480-290.mis2
<400> 567
                                                                        19
aaaaggatca gtggttgcc
<210> 568
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> microsequencing oligo for 99-1481-285.mis2
<400> 568
                                                                        19
taccatcttg aggttagag
<210> 569
<211> 19
<212> DNA
<213> Homo Sapiens
<220>
<221> primer_bind
<222> 1..19
<223> potential microsequencing oligo for 99-1484-101.mis2
<400> 569
                                                                        19
agattttaag gagaggagt
<210> 570
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> potential microsequencing oligo for 99-1484-328.mis2
 <400> 570
                                                                         19
 tctgaaaact gaatccctt
 <210> 571
 <211> 19
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> primer_bind
 <222> 1..19
 <223> microsequencing oligo for 99-1485-251.mis2
 <400> 571
                                                                         19
 aggggacatt cttggttct
 <210> 572
```

<211>	19	
<212>	DNA .	
<213>	Homo Sapiens	
<220>		
<221>	primer_bind	
<222>	119	
<223>	potential microsequencing oligo for 99-1490-381.mis2	
<400>		
agatgo	acag tagcgtacc	19
<210>		
<211>		
<212>	DNA	
	Homo Sapiens	
<220>		
	primer_bind	
<222>		
<223>	microsequencing oligo for 99-1493-280.mis2	
<400>		
	cagec aaaccccat	19
<210>		
<211>		
<212>		
	Homo Sapiens .	
<220>	nomo Edpieno	
	primer_bind	
<222>		
	microsequencing oligo for 99-151-94.mis2	
<400>		
	gattt tgaaatttta gaa	23
<210>		
<211>		
<211>		
	Homo Sapiens	
<220>	HOMO Sapiens	
	number hind	
	primer_bind 119	
<2222 <2222	potential microsequencing oligo for 99-211-291.mis2	
<400>		
		19
_	acctg ttgaaaaca	
<210><211>		
<211>		
	Homo Sapiens	
<220>		
	primer_bind	
	119 :	
~222×	potential microsequencing oligo for 99-213-37.mis2	
<400>		
	cactg gagtcctcc	19
<210>		
<211>		
<212>		
	Homo Sapiens	
<220>		
	nrimer hind	
	2119	
	potential microsequencing oligo for 99-221-442.mis2	
<400>		
	ggctag gtcatggaa	19
<210>		
<211>		

PCT/IB98/02133

217

<212> <213>	DNA Homo Sapiens
<220>	
<221>	primer_bind
<222>	
<223>	potential microsequencing oligo for 99-222-109.mis2
<400>	578
ctgaag	gaaat tcatatcgt

INTERNATIONAL S: \RCH REPORT

Inter Application No PCT/1B 98/02133

(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category '	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EMBL Database Entry HS1184481 Accession number AA280082;20 June 1997 National Cancer Institute, Cancer Genome Anatomy Project (CGAP): "zs93h09.r1 NCI_CGAP_GCB1 Homosapiens cDNA clone IMAGE: 705089 5'."100% overlap in 170bp p with Seq ID Nos 121 and 69; & 100% overlap in 237bp with ID No 112." XP002109142 "100% overlap in 170bp with Seq ID no 3." Additionally, an 100% overlap in 100 and 101 amino acids with polypeptides having Seq ID Nos 135 and 134 respectively."	1-3,5-9, 16-18
A	ICHIKAWA T ET AL: "Metastasis supressor genes for prostate cancer" PROSTATE, vol. 28, no. 3 suppl 6, 1966, pages 31-35, XP002109139 the whole document	1-3,5-38
A	G S BOVA ET AL: "Homozygous deletion and frequent allelic loss of cchromosome 8p22 loci in human prostate cancer" CANCER RESEARCH, vol. 53, 1 September 1993 (1993-09-01), pages 3869-3873, XP002084748 ISSN: 0008-5472 the whole document	1-3,5-38
A	WU C ET AL: "Deletion mapping defines three discrete areas of allelic imbalance on chromosome arm 8p in oral and oropharangeal squamous cell carcinomas" GENES, CHROMOSOMES AND CANCER, vol. 20, no. 4, December 1997 (1997–12), pages 347–53, XP002109140 page 352, paragraph 3	1-3,5-38
Α	WO 96 20288 A (CTRC RES FOUNDATION ;UNIV MICHIGAN (US)) 4 July 1996 (1996-07-04) the whole document	1-3,5-38
Α	WO 97 46702 A (UNIV CALIFORNIA) 11 December 1997 (1997-12-11) the whole document	1-3,5-38
Α	WO 97 36535 A (UNIV TEXAS) 9 October 1997 (1997-10-09) the whole document/	1-3,5-38

PCT

ORLD INTELLECTUAL PROPERTY ORGANIZAT International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: C12N 15/85, 9/10, C07K 16/30, 16/40, C12Q 1/68, A01K 67/027, C07K 14/47

(11) International Publication Number:

WO 99/32644

(43) International Publication Date:

1 July 1999 (01.07.99)

(21) International Application Number:

PCT/IB98/02133

A3

(22) International Filing Date:

2? December 1998 (22.12.98)

(30) Priority Data:

08/996,306 60/099,658

22 December 1997 (22.12.97) US 9 September 1998 (09.09.98) US

(71) Applicant (for all designated States except US): GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR).

(72) Inventors; and

- (75) Inventors/Applicants (for US only): COHEN, Daniel [FR/FR]; 5, avenue Odette, F-94210 Fontenay-sous-Bois (FR). BLUMENFELD, Marta [FR/FR]; 5, rue Tagore, F-75013 Paris (FR). CHUMAKOV, Ilya [FR/FR]; 196, rue des Chèvrefeuilles, F-77000 Vaux-le-Pénil (FR). BOUGUEL-ERET, Lydie [FR/FR]; 108, avenue Victor-Hugo, F-92170 Vanves (FR).
- (74) Agents: MARTIN, Jean-Jacques et al.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report: 10 September 1999 (10.09.99)

(54) Title: PROSTATE CANCER GENE

(57) Abstract

The present invention relates to PG1, a gene associated with prostate cancer. The invention provides polynucleotides including biallelic markers derived from PG1 and from flanking genomic regions. Primers hybridizing to these biallelic markers and regions flanking are also provided. This invention provides polynucleotides and methods suitable for genotyping a nucleic acid containing sample for one or more biallelic markers of the invention. Further, the invention provides methods to detect a statistical correlation between a biallelic marker allele and prostate cancer and between a haplotype and prostate cancer. The invention also relates to diagnostic methods of determining whether an individual is at risk for developing prostate cancer, and whether an individual suffers from prostate concer as a result of a mutation in the PG1 gene.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain		LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland		LT	Lithuania	SK	Slovakia
AT	Austria	FR	France		LU	Luxembourg	SN	Senegal
ΑÜ	Australia	GA	Gabon		LV	Latvia	SZ	Swaziland
ΑZ	Azerbaijan	GB	United Kingdom		MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia		MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana		MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea		MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary		ML	Mali	TT	Trinidad and Tobago
ВJ	Benin	IE	Ireland		MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel		MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland		MW	Malawi	US	United States of America
CA	Canada	IT	Italy		MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	•	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	1	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Кугдуzstan		NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's		NZ	New Zealand		
CM	Cameroon		Republic of Korea		PL	Poland		
CN	China	KR	Republic of Korea		PT	Portugal		
CU	Cuba	KZ	Kazakstan		RO	Romania		
CZ	Czech Republic	LC	Saint Lucia		RU	Russian Federation		
DE	Germany	LI	Liechtenstein		SD	Sudan		
DK	Denmark	LK	Sri Lanka		SE	Sweden		
EE	Estonia	LR	Liberia		SG	Singapore		

${\bf INTERNATIONAL}_{\overbrace{}{}}{\stackrel{\textstyle <}{=}}{{\bf ARCH}}\,{\bf REPORT}$

Ir inal Application No PCI/IB 98/02133

a. classi IPC 6	FICATION OF SUBJECT MATTER C12N15/85 C12N9/10 A01K67/027 C07K14/47	C07K16/30	C07K16/40	C12Q1/68
According to	International Patent Classification (IPC) or to both	national classification a	and IPC	
	SEARCHED			
Minimum do IPC 6	cumentation searched (classification system follow C12Q C07K C12N A01K	ved by classification sy	nbols)	
Documentat	ion searched other than minimum documentation t	o the extent that such d	ocuments are included in	the fields searched
Electronic d	ata base consulted during the international search	(name of data base an	d. where practical, search	terms used)
C. DOCUM	ENTS CONSIDERED TO BE RELEVANT			
Category °	Citation of document, with indication, where app	ropriate, of the relevant	passages	Relevant to claim No.
X	EMBL Database Entry HS10 Accession Number H06164 Hillier L et al.: "y1770 cDNA clone 44264 5'." 100% identity in 169pb No 118. XP002109141 see database entry	; 23 June 199 g12.r1 Homosa	apiens Seq ID	1-3
X Furt	her documents are listed in the continuation of box	(υ. <u>Χ</u>	Patent family member	ers are listed in annex.
"A" docum consider filling of the country country of the country o	ent defining the general state of the art which is not defend to be of particular relevance document but published on or after the internation date ent which may throw doubts on priority claim(s) or is cited to establish the publication date of another on or other special reason (as specified) ent referring to an oral disclosure, use, exhibition means lent published prior to the international filling date be than the priority date claimed	al -X- -Y- or ,	or priority date and not in cited to understand the p invention document of particular rele cannot be considered no involve an inventive step document of particular rele cannot be considered to document is combined w	after the international filing date conflict with the application but rinciple or theory underlying the evance; the claimed invention well or cannot be considered to when the document is taken alone evance; the claimed invention involve an inventive step when the lith one or more other such docubeing obvious to a person skilled
Date of the	actual completion of the international search		Date of mailing of the inte	ernational search report
1	15 July 1999		29/07/1999	
Name and	mailing address of the ISA European Patent Office, P.B. 5818 Patentlaa NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fay: (-31-70) 340-3016	n 2	Authorized officer Osborne H	

Form PCT/ISA/210 (second sheet) (July 1992)

${\bf INTERNATIONAL} {\color{red} \leftarrow} {\bf EARCH} \ {\bf REPORT}$

Ir anal Application No PCT/IB 98/02133

		PCT/IB 98/02133
C.(Continu	ation) DOCUMENTS CONSIDERED TO BE RELEVANT	
Category ·	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KRUGLYAK L: "THE USE OF A GENETIC MAP OF BIALLELIC MARKERS IN LINKAGE STUDIES" NATURE GENETICS, vol. 17, no. 1, 1 September 1997 (1997-09-01), pages 22-24, XP002050647 ISSN: 1061-4036 the whole document	19-38
А	WANG D ET AL: "TOWARD A THIRD GENERATION GENETIC MAP OF THE HUMAN GENOME BASED ON BI-ALLELIC POLYMORPHISMS" AMERICAN JOURNAL OF HUMAN GENETICS, vol. 59, no. 4, 1 October 1996 (1996-10-01), page A03 XP002050641 ISSN: 0002-9297 the whole document	19-38
A .	SCHORK N J ET AL: "Linkage disequilibrium mapping for quantitative traits within case/control settings" AMERICAN JOURNAL OF HUMAN GENETICS, vol. 61, no. 4, SUPPL, 1 January 1997 (1997-01-01), page A293 XP002089399 ISSN: 0002-9297	19-38
	500	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB 98/02133

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
Claims Nos.: 4, 39 because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
This International Searching Authority found multiple inventions in this international application, as follows:
As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report
covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it, is covered by claims Nos.:
The additional search fees were accompanied by the applicant's protest. No protest accompanied the payment of additional search fees.

INTERNATIONAL * EARCH REPORT

International Application No. PCT/IB 98 02133

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 4 and 39

Present claims 4 and 39 relate to an extremely large number of possible polynucleotide sequences. In the present case, a meaningful search over the whole of the claimed scope is impossible. Consequently, the search has been carried out for those parts of the claims which appear to be related to the PG1 gene per se., and their utility in the claimed diagnostic methods.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL ST RCH REPORT

Informs - on patent family members

Inter Application No
PCT/1B 98/02133

Patent document cited in search report		Publication date	Patent family member(s)	Publication date			
WO 9620288	Α	04-07-1996	US 5658730 A AU 4473596 A	19-08-1997 19-07-1996			
WO 9746702	Α	11-12-1997	NONE				
WO 9736535	А	09-10-1997	US 5861248 A AU 2601797 A EP 0904405 A	19-01-1999 22-10-1997 31-03-1999			

cagtgtcact aagcctgtag ttcactgatg aagctagagt gcccactagt gagccactat 33240 cactagtgag cacactagta atagtgtgaa agaaagtgac ttccactgtc ccatgaacac 33300 acceaectte ateteeteta cettgeaect ggtteagatt teggeagatg cagtgageat 33360 ggtgcttaag aagtctgaag gatagctctg ggggatggtg gcacatgaag gtcacatctt 33420 tttttaatta ggtgatttaa atgggtttgt gcacatgagt gcaacaccca tacaggccag 33480 aagggggcat aagateeeeg getetagagt gteaggetgt tgtgaactge etggaatagg 33540 tgttgggaac tgatcttggg tcccctggaa gagcattgag tacatgtcat cactatctct 33600 ccagcctcac actcttatcg gtatgtctca tttgtggggg caatttgggg agtcccacct 33660 ctggatcagg tactgtgaaa tatggagttg gcggggctgg atccctgtct tgccaccagc cagctgagga aagctgtcaa ttgtcttcct gtgtctgcct cagtttccta gaaactagaa 33780 aggaaaaatg gatggtatca ctaagttcag ccttccattg taaggatcca ttgaagtagt 33840 tggtgtgatg tactcacgct ggtgcccctc cccttctgag ctgcaagcat cagctgttgg 33900 acccagcagt totgtgotoc gacaggaago agtgggaaac tgggotgcaa gatgatocca 33960 gttctagcat ttgctgcaac cccctttgct cacatctgtt tccacatttt atttcatcct 34020 tgagcacata accttttcat tttgatacat gcttttcttt acatagcctg ggccagcctt 34080 gaactcatgc teeteetgee ataatteett aaggaetgae aatgaeagtg gaeatgeaec 34140 gccaggtccc actgctccta gtacttttgt tatgggtctt ctatgctggt ctaatgtgga atgtgacact gcacaccagg cattgtggga tgaagtagaa catgttcatg cacacaaaga 34200 34260 tcaatcccaa acagcatctg cccacccctc ccctctgccc cttcccccgg tggatgttag 34320 cattetettg tgtagtgetg ggteggeece ttetetgtet etaaacattg aaacaagggg 34380 aacagaccca tacataactc caacacagcg atctggtcca agtctggatg tagaaccctt 34440 ggetetgete tttteteete tteecagage tgteecagte geeteeettt etaattgget 34500 ggtgctcatc taacttgatg tatatgtttc tttcctggtc tgttttatga ctggcctgct 34560 ccagtcatta gtgcatctgg tgttagaagc taagctcaac ttggcctcac agtcttgatg 34620 ttcaggacac atggggttat ggctggatcc tgtgtcaagc acatcacttc tttctgaacc 34680 cacacatett aagacatggg cattgteaca tggetgacag cagtacatte ttgtatgtag ttttetetea agtgttttgg ttacatggee ctaaageeta ggaetgtetg tetteaatea 34740 34800 tgtctgccac ctgctgccca gcaagcccaa gttgtggttc ttcctgtcta atctgcctct 34860 tttatcttta gccctcctcc ataagcctct tttccactga gctctgggtt attcagatta 34920 cageteecta etgetteet gagaceetaa geegeteeae gatgeacage ageaettee 34980 agcagtccta tggagatgca gtgtcgagat gatatgagct gtgctgtata tgtaaaatgc 35040 atggtagaat ctgaaaacca agcacaaaaa aataaaaaaa aagttcatta atatttatag 35100 caatcatgta atgttttgaa aatacgtgcc taataaaatt attttcacct ttttaaaaat 35160 gtgggtacta tgatgtttaa tagcgcatat tcatgtggca tatttacctc aatggggcat 35220 ggaaaacaag acgacaaagg gtgggatatt tctctcaact ttcctagctg caacttgtga 35280 taatttgtga gacacattaa gcactcaaga agggtttgat tctgtagaca aaacccgggg 35340 tcagacagtc cacagcttca gcatagtgtg ttcactatat ctaaagtagg ttttacttct 35400 ttaagattaa ttttgttctt agacagtttc ccacatgcag cacaaagcag ttatgtttta 35460 attttatagg ctacaatgac ggtatttcaa aaagtaagat gtgggagtta caccttgaat 35520 gatcacccaa ggattgttgg ccatcagcac tgaatggccc agaaccaatc tacagcaaat 35580 ctgagagage agagacatge tgetecactg tetgagggee ettgeceatg gtggttgeta 35640 ttgacatctt gacaggatct aaaatcacct aggtgtgctc tgggcacctc tggacatatc tgggagtete tagactggat tgcctgaggt tgatttaaaa accetaactg ggcagcacca 35760 ttcattggga tggggtcctg gcgtggataa agattggagg atgagggagc accaggttcc 35820 acctetetge etectgactg gatgeageat gateagetee eteteetget gtgacacaat 35880 ctctgtaccc tcaaactgaa caaaacaagc tcactcgaaa gttgctttcg ttagtcactt 35940 tgtcacagct ttgtcacagt gatgagaaaa ataacacaca ccccccaaat gaggattgga 36000 tegaggeest gggeatgetg ggtaaactet teaceactaa getatattee taaceetttt 36060 tactttgaat agggtcttaa ttgcccagat tggccttaaa atgataattc tcctgtcttc 36120 agcttcccaa gcagctggaa ttacaaatgc gagcactagt atgtatatgt agtatataaa taatatattg tacatatatt cacatgtgtg tgtacatgtg tgtacacaat tacaaatggg 36240 agccccagaa tgatgtatac acatacatat aagtaggtaa ttaatatata catatttaca 36300 catattatac tagtgtgtat atatagaata tataatatgt attgcacata tattcacata 36360 tgtgtgtaca tgtccatgtc acattgtgat tgctagttaa ataccacttt tccctgctct 36420 ctaatccaag tagcaattga acctgtgaat tatggtaaac ttcaggggat tgaaaagcct 36480 gacctccaca gaatcagcta acgttagctg ctactatgat tccaagcagt agttactgga 36540 agtgtttgct cctgtattct cttggaagga agaactagat tgttagtcat cagttcacag 36600 toggttatgg tococotgtt ggcotcatag accagaaagt agagttgcat ggtttgattg taaagtgacc cacctcccac aggttcacat gcttgaacac tgggtcccca gatgatggag 36720 ctcttggggg aggttgtgga gaccatctag gaggttagct tacctggagg aaatgggctt 36780 ctaggggaca gacattgaag atgataggtt ttccctgttt ctgccctggg aatatatcca 36840

cctca gaggg tggga tgtca aagaat ctca ctcc atgtct agttt agttca ctc2 2211 <2112 <213 <410	aaacccaaacccaaacccgggggggggggggggggggg	cc to	gaged cegage aggga cetgg gaaca tegaa ceatga acagt tated aaaag gaaca gaaca seul	agag gcaca gctgt laaca gtgtt latag ggtt lataca ggttaca gtaca ggtac gtaca ggtac gtaca ggtac gtaca ggtac ggta ggta	g tag g g g g g cag t cag t cag t cag t cag t cag t c c c c c c c g g g g a t c c a a	yatco atcat ycctt atggg atcaa gtgct atcaa cctaa actcaa actcaa actcaa acca acca ac ac	ette etgt ggag ggat aagt cett caat cetet acag cgca ttca acag ggag gag gag gag gag gag gag gag	gtcortget ggarget tear aggreet tear gtcort tear gtcort tear aggreet tear at the cortan	ctttgcattcattcattcattagcattcattcatcatcatcatcatcatcatcatcatcattcattagcattcatt	get a	agatocogatoc	yttti	gaat acct ggtcttaaca	cacggactgatgaaggacccaggacctcacccaggacaccccaccccaggacacccccccc	tctac gcagt gagcct tcagg ttcctg gtcctat gagcccattcag ggcactt tcagg gccattcag agagg tctagaga atgct gatgct gatgct atgct gatgct atgct	36900 36960 37020 37080 37140 37200 37260 37320 37380 37560 37560 37620 37680 37740 37800 37860 37950
gago	cgag	ag g	atgc	tgct	g tc	cctg	gtgc	tcc	acac	gta	ctct	atg Met 1	cgc Arg	tac Tyr	ctg Leu	56
Leu	ccc Pro	agc Ser	gtc Val	ctg Leu	ttg Leu 10	ctg Leu	ggc Gly	tcg Ser	gcg Ala	ccc Pro 15	acc Thr	tac	ctg Leu	ctg Leu	gcc Ala 20	104
5 tgg Trp	acg Thr	ctg Leu	Trp	cgg Arg 25	ata	ctc Leu	tcc Ser	gcg Ala	ctg Leu 30	atg	ccc Pro	gcc Ala	cgc Arg	ctg Leu 35	tac Tyr	152
cag Gln	cgc Arg	gtg Val	gac	gac	cgg Arg	ctt Leu	tac Tyr	tgc Cys 45	gtc Val	tac Tyr	cag Gln	aac Asn	atg Met 50	gtg Val	ctc Leu	200
ttc Phe	ttc Phe	ttc Phe 55	gag	aac Asn	tac Tyr	acc Thr	ggg Gly 60	gtc Val	cag Gln	ata Ile	ttg Leu	cta Leu 65	tat Tyr	gga Gly	gat Asp	248
ttg Leu	cca Pro 70	aaa	aat Asn	aaa Lys	gaa Glu	aat Asn 75	gta Val	ata Ile	tat Tyr	cta Leu	gcg Ala 80	aat Asn	cat His	caa Gln	agc Ser	296
aca Thr 85	att	gac Asp	tgg Trp	Ile	gtt Val 90	gcg Ala	Asp	Met	Leu	Ala	gcc Ala	Arg	Gin	Asp	Ala	344
cta	gga Gly	cat His	gtg Val	cac	tac	gta Val	ctg Leu	aaa Lys	gac Asp 110	aag Lys	tta Leu	aaa Lys	tgg Trp	ctt Leu 115	ccg Pro	392
ctg Leu	tat Tyr	Gly ggg	ttc Phe 120	tac	ttt Phe	gct Ala	cag Gln	cat His 125	gga Gly	gga Gly	att Ile	tat Tyr	gta Val 130	aaa Lys	cga Arg	440
agt Ser	gcc Ala	Lys	ttt	aat Asn	gat Asp	aaa Lys	gaa Glu 140	atg Met	aga	agc Ser	aag Lys	ctg Leu 145	cag Gln	agc Ser	tat Tyr	488
gtg Val	Asn	Ala	gga Gly	aca Thr	ccg Pro	atg Met 155	tat Tyr	ċtt	gtg Val	att Ile	ttc Phe 160	cca	gag Glu	gga Gly	aca Thr	536
Arg	Туг	aat	gca Ala	aca Thr	tac Tyr 170	aca Thr	aaa	ctc Leu	ctt Leu	tca Ser 175	gcc Ala	agt Ser	cag Gln	gca Ala	ttt Phe 180	584
165 gct	gct	cag	cgg	ggc	ctt	gca	gta	tta	aaa			ctg	aca	. cca	aga	632

Ala	Ala	Gln	Arg	Gly	Leu	Ala	Val	Leu	Lys	His	Val	Leu	Thr	Pro	Arg	
				185 cac					190					195		680
Ile	Lys	Ala	Thr	His	Val	Ala	Phe	Asp 205	Ser	Met	ьуs	ser	11S	Leu	Asp	
Ala	Ile	Tyr 215	Asp	gtc Val	Thr	Val	Val 220	Tyr	GIu	GIY	Asn	225	гуѕ	GIY	ser	728
Gly	Lys	tac Tyr	Ser	aat Asn	Pro	Pro 235	Ser	Met	Thr	Glu	Phe 240	Leu	Cys	гуѕ	GIN	776
Cys 245	cca Pro	Lys	Leu	cat His	Ile 250	His	Phe	Asp	Arg	11e 255	Asp	Arg	Asn	GIU	260	824
CCA	gag Glu	gaa Glu	caa Gln	gaa Glu 265	cac	atg Met	aaa Lys	aag Lys	tgg Trp 270	ctt Leu	cat His	gag Glu	cgc Arg	ttt Phe 275	gag Glu	872
ata Ile	aaa Lys	gat Asp	Arg	ttg Leu	ctc Leu	ata Ile	gag Glu	ttc Phe 285	tat Tyr	gat Asp	tca Ser	cca Pro	gat Asp 290	cca Pro	gaa Glu	920
aga Arg	aga Arg	Asn	280 aaa Lys	ttt Phe	cct Pro	ggg Gly	aaa Lys 300	agt	gtt Val	cat His	tcc Ser	aga Arg 305	cta Leu	agt Ser	gtg Val	968
aag Lys	aag Lys 310	Thr	tta Leu	cct	tca Ser	gtg Val 315	tta	atc Ile	ttg Leu	ggg Gly	agt Ser 320	ttg Leu	act Thr	gcg Ala	gtc Val	1016
Met	ctg Leu	atg Met	Thr	Glu	Ser 330	gga Gly	Arg	Lys	Leu	Туr 335	Met	GIY	Thr	Trp	ttg Leu 340	1064
tat	gga Gly	acc Thr	cto Leu	ctt Leu 345	Gly	tgc Cys	ctg Leu	tgg Trp	ttt Phe 350	val	att Ile	aaa Lys	gca Ala	taa * 355		1109
tcctgaatta aataaggagt tttcttgttg ttgttttttt tgttttgttt												1169 1229 1289 1349 1381				
<2: <2:	22> 23>	1 2 5	47 leme	nt p										107.	mis2	
ct <2 <2 <2 <2	aaac 10> 11> 12>	aacc 186 47 DNA		aaat		taca	gcaa	cc a	ggca	aatg	c ct	gata	g			47

```
<221> allele
<222> 1..47
<223> polymorphic fragment 4-14-317
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-14-317.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-14-317.mis2
<400> 186
                                                                        47
cataacatgc aaggtgggca agaaaaagag gtgggcacag ctcatga
<210> 187
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-14-35
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-14-35.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-14-35.mis2
                                                                        47
atccaacaca gaaaccgcta aaaccaggca gaagctgtct gcagaga
<210> 188
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
 <223> polymorphic fragment 4-20-149
 <221> allele
 <222> 24
 <223> polymorphic base C
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-20-149.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-20-149.mis2
                                                                         47
 tttttgctgt gtcttcaaag tgactcttgg tttattgcct gctaagg
 <210> 189
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-20-77
 <221> allele
```

PCT/IB98/02133 WO 99/32644

```
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-20-77.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-20-77.mis2
<400> 189
                                                                        47
tgcaacatga agattctgaa gggactttgt tgtctgagaa cacatct
<210> 190
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-22-174
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-22-174.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-22-174.mis2
<400> 190
                                                                        47
ggattgtgca gaagttgcct ttcatgttca aaaatgttaa tttgttt
<210> 191
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-22-176
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-22-176.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-22-176.mis2
 <400> 191
                                                                         47
attgtgcaga agttgccttt catattcaaa aatgttaatt tgtttgt
 <210> 192
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-26-60
 <221> allele
 <222> 24
 <223> polymorphic base A
 <221> primer_bind
 <222> 1..23
```

```
<223> potential microsequencing oligo 4-26-60.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-26-60.mis2
                                                                       47
gatgggaaag tgcatcttaa gacagttagc aggccaagga gcgactt
<210> 193
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-26-72
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-26-72.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-26-72.mis2
<400> 193
                                                                        47
catcttaaga cagttagcag gccaaggagc gactttaaag ggtgagc
<210> 194
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-3-130
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-3-130.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-3-130.mis2
 <400> 194
                                                                        47
 tattgggcct aaaacagtat tctataaagc ttaaattggt attaact
 <210> 195
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-38-63
 <221> allele
 <222> 24
 <223> polymorphic base A
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-38-63.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-38-63.mis2
```

```
<400> 195
                                                                       47
tataagttat aagaaaatca ggcagaggct aaactttttt tttgttt
<210> 196
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-38-83
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-38-83.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-38-83.mis2
<400> 196
ggcagaggct aaactttttt tttgtttggc aatgctgttg agaatat
                                                                        47
<210> 197
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-4-152
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-4-152.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-4-152.mis2
<400> 197
tactttccca ttgttcctga cttcgttatc ctatatataa acagaaa
                                                                        47
<210> 198
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-4-187
 <221> allele
 <222> 24
 <223> polymorphic base A
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-4-187.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-4-187.mis2
 <400> 198
                                                                         47
 tataaacaga aacatggatg agtaaaaaaa aaaaaaaaa aaaaaaa
 <210> 199
 <211> 47
```

WO 99/32644 PCT/IB98/02133 134

```
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-4-288
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-4-288.misl
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-4-288.mis2
<400> 199
                                                                       47
ctgtcatcaa ctaattttca caagtaccta tgttttgatt tcatgta
<210> 200
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-42-304
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-42-304.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-42-304.mis2
<400> 200
                                                                        47
attatttaaa actatttatg taaccttatt ttcaggggtt tttaatt
<210> 201
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
 <222> 1..47
<223> polymorphic fragment 4-42-401
 <221> allele
 <222> 24
 <223> polymorphic base A
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-42-401.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-42-401.mis2
                                                                        47
 taagaaagaa ttctgtgttc tggacaaagt ttaaacccac agagcca
 <210> 202
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
```

```
<222> 1..47
<223> polymorphic fragment 4-43-328
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-43-328.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-43-328.mis2
<400> 202
                                                                         47
agaattctgt gttctggcca aagcttaaac ccacagagcc agtttaa
<210> 203
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-43-70
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind <222> 1..23
<223> potential microsequencing oligo 4-43-70.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-43-70.mis2
<400> 203
                                                                         47
atcgcctcca ttattctcaa aaagaccatg ggacacaaca caagaag
<210> 204
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-209
<221> allele
 <222> 24
 <223> polymorphic base C
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-50-209.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-50-209.mis2
 <400> 204
 atatagagtg tgcatccctg acactgaaac tgaaggcttt atggttt
                                                                          47
 <210> 205
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-50-293
 <221> allele
 <222> 24
```

WO 99/32644

```
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-293.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-293.mis2
<400> 205
                                                                        47
cctgagtccc agggggctga caggggacag tttaaaacat tgatgaa
<210> 206
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-323
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-50-323.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-323.mis2
<400> 206
                                                                        47
tttaaaacat tgatgaatct ttactactac aaaagggttc gatttag
<210> 207
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-50-329
<221> allele
<222> 24
<223> polymorphic base C
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-50-329.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-50-329.mis2
 <400> 207
                                                                         47
 acattgatga atctttatta ctacaaaagg gttcgattta ggctagc
 <210> 208
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-50-330
 <221> allele
 <222> 24
 <223> polymorphic base A
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-50-330.mis1
```

PCT/IB98/02133

```
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-50-330.mis2
<400> 208
cattgatgaa totttattac tacaaaaggg ttcgatttag gctagcc
                                                                        47
<210> 209
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-52-163
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-52-163.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-52-163.mis2
<400> 209
                                                                        47
gaacaggata ttcttaacta ccaaagaatt ttacacatct attgttt
<210> 210
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-52-88
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-52-88.mis1
<221> primer_bind <222> 25..47
<223> complement potential microsequencing oligo 4-52-88.mis2
<400> 210
                                                                         47
tccatgtcat tattattcaa aagcttaaaa aatacacaag gtgaaaa
<210> 211
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
 <222> 1..47
 <223> polymorphic fragment 4-53-258
 <221> allele
 <222> 24
 <223> polymorphic base A
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-53-258.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-53-258.mis2
 <400> 211
```

```
47
gagaaatcat gcagagagaa tgcattctca ctcaaatttt aacctaa
<210> 212
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-54-283
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-54-283.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-54-283.mis2
<400> 212
                                                                       47
aagtagtttt tcacactttc tctatgatac aatcgatggc ttaatct
<210> 213
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-54-388
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-54-388.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-54-388.mis2
<400> 213
                                                                        47
ctctctatcg tatacatctt tacacacgct gcagcgccaa gactcca
<210> 214
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-55-70
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-55-70.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-55-70.mis2
 <400> 214
                                                                        47
 tattaagaac ctaggtttta aaaaactctc tatcgtatac atcttta
 <210> 215
 <211> 47
 <212> DNA
```

PCT/IB98/02133

```
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-55-95
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-55-95.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-55-95.mis2
<400> 215
                                                                        47
ctctctatcg tatacatctt tacacacgct gcagcgccaa gactcca
<210> 216
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-56-159
<221> allele <222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-56-159.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-56-159.mis2
<400> 216
                                                                         47
aagttttcct tctcttctgt agacgtctcc atgttacagt caactat
<210> 217
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
 <223> polymorphic fragment 4-56-213
 <221> allele
 <222> 24
 <223> polymorphic base A
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-56-213.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-56-213.mis2
 <400> 217
                                                                         47
 atggctcatg ttcactctgg ttcaccttca qaggagtttg atatttt
 <210> 218
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
```

WO 99/32644

```
<223> polymorphic fragment 4-58-289
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-53-289.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-58-289.mis2
<400> 218
                                                                        47
catacctgca gcctgctttt ggtgaggggt gactacttta cctgcaa
<210> 219
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-58-318
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-58-318.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-58-318.mis2
tgactacttt acctgcaata tttatttgca agtttatttc ttccttt
                                                                        47
<210> 220
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-60-266
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-60-266.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-60-266.mis2
 aacaggacca agacactgca ttagataaag tttcagtatt tcttagc
                                                                         47
 <210> 221
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 4-60-293
 <221> allele
 <222> 24
 <223> polymorphic base C
```

```
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-60-293.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-60-293.mis2
<400> 221
aagtttcagt atttcttagc agacgaagcc agcaggaagt cctccta
                                                                         47
<210> 222
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-84-241
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-84-241.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-84-241.mis2
gaaaaaaaaa tagtgactgc cacggtgaat aattcagttc itcagaa
                                                                         47
<210> 223
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-84-262
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-84-262.mis1
<221> primer_bind <222> 25..47
<223> complement potential microsequencing oligo 4-84-262.mis2
<400> 223
                                                                         47
acggtgaata attcagttct tcaaaagcag caacatgatc tcatgga
<210> 224
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-86-206
<221> allele <222> 24
<223> polymorphic base A
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-86-206.mis1
 <221> primer_bind
```

```
<222> 25..47
<223> complement potential microsequencing oligo 4-86-206.mis2
                                                                       47
gtattcaaat caggacacac cacaaatggc atctacacgt taacatt
<210> 225
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-86-309
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-86-309.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-86-309.mis2
<400> 225
                                                                       47
tggctctagg caggccactt tagagagtga ggaaccagag agcagaa
<210> 226
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-88-349
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-88-349.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-88-349.mis2
<400> 226
gaaactaaaa gacaatatto agtgtgagat tttccaagtt ctttatg
                                                                        47
<210> 227
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
 <223> polymorphic fragment 4-89-87
 <221> allele
 <222> 24
 <223> polymorphic base C
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 4-89-87.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 4-89-87.mis2
 <400> 227
                                                                        47
 ttcttccctg aacgctggtt tcacatagtt tttgtgttga gaataga
```

```
<210> 228
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-123-184
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-123-184.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-123-184.mis2
<400> 228
                                                                        47
ccagcccaga acattcacca gctgggccaa gagttctgct gggtttt
<210> 229
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-128-202
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-128-202.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-128-202.mis2
<400> 229
                                                                         47
aatgtctgtt tcttagagaa ctgaaacaca cacacataca tacacac
<210> 230
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-128-275
 <221> allele
 <222> 24
 <223> polymorphic base A
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-128-275.mis1
 <221> primer_bind <222> 25..47
 <223> complement potential microsequencing oligo 99-128-275.mis2
                                                                         47
 acacccctac ctcacatgtg tagacaaatg tatgcatata tgtctct
 <210> 231
 <211> 47
 <212> DNA
 <213> Homo Sapiens
```

```
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-128-313
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-128-313.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-123-313.mis2
                                                                        47
tatgtctcta gacagatata cataagattc tatttggcat agaaaaa
<210> 232
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-128-60
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-128-60.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-128-60.mis2
<400> 232
                                                                        47
gcactgtgac ccaggcgcta ggtccctctt acagtgacac tccgaca
<210> 233
<211> 47
<212> DNA
<213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-12907-295
 <221> allele
 <222> 24
 <223> polymorphic base A
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-12907-295.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-12907-295.mis2
 <400> 233
                                                                         47
 gctatatggc attatatctc cacagggcag acctgatgta caagatg
 <210> 234
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-130-58
```

```
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-130-58.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-130-58.mis2
<400> 234
                                                                        47
aaagcaaaag agcttcaaaa atacttcagg agtgtgcata tggcgag
<210> 235
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-134-362
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-134-362.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-134-362.mis2
<400> 235
                                                                        47
caaaacactc atgttagtta gatgattatt cctattacaa agataag
<210> 236
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-140-130
<221> allele
 <222> 24
<223> polymorphic base C
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-140-130.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-140-130.mis2
 <400> 236
                                                                         47
 tgttcaaaag cagctacaga ccacatgtaa acaattgagc atggctg
 <210> 237
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-1462-238
 <221> allele
 <222> 24
 <223> polymorphic base G
 <221> primer_bind
```

```
<222> 1..23
<223> potential microsequencing oligo 99-1462-238.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1462-238.mis2
<400> 237
                                                                       47
ccctttcaag gttagtaact catgtgctgt gtttctgctt cagaagg
<210> 238
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-147-181
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-147-181.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-147-181.mis2
<400> 238
gtgtcatgaa aaagagcatg ataaaaagaa aaacttaaat ctttata
                                                                       47
<210> 239
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1474-156
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1474-156.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1474-156.mis2
<400> 239
                                                                       47
cttgtactca taagttaaat attgataaca agaagaaata tggactt
<210> 240
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1474-359
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1474-359.mis1
<221> primer_bind
<222> 25..47
```

```
<223> complement potential microsequencing oligo 99-1474-359.mis2
aaaaaaaatc aaattattgt accaaattcc ctaatatcag atgtgta
                                                                       47
<210> 241
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1479-158
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1479-158.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1479-158.mis2
<400> 241
                                                                       47
tttaaaaatc cacttgtaat cgccgctaat tggagtgtat attcagg
<210> 242
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1479-379
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1479-379.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1479-379.mis2
<400> 242
gtagagctgt gtactgaggt cagagaagca gctcatggta cagcctt
                                                                       47
<210> 243
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-129
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-129.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-129.mis2
<400> 243
ttcatatcta tacaaataat tttaaattta atacataggg ctgcaaa
                                                                       47
<210> 244
```

```
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-132
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-132.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-132.mis2
<400> 244
                                                                       47
atatctatac aaataatttt gaacttaata catagggctg caaaaca
<210> 245
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-139
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-139.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-139.mis2
<400> 245
                                                                        47
tacaaataat tttgaattta atacataggg ctgcaaaaca aggttga
<210> 246
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-140
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-140.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-140.mis2
<400> 246
acaaataatt ttgaatttaa tacatagggc tgcaaaacaa ggttgat
                                                                        47
<210> 247
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
```

```
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-182
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-182.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-182.mis2
<400> 247
                                                                        47
ttgatgttga tatgggcaac tgtatgttgg atggtcccaa agcattc
<210> 248
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-366
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-366.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-366.mis2
                                                                        47
teettgteaa aggtetetee etggtgetea eggetgeege etcaaag
<210> 249
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-148-76
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-148-76.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-148-76.mis2
<400> 249
                                                                        47
tgatagaatg ccttcctgaa ttactactct tgatggcttc ataaaac
<210> 250
<211> 47
<212> DNA
<213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-1480-290
 <221> allele
```

```
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1480-290.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1480-290.mis2
<400> 250
                                                                       47
tgcaccatct tcaccacaac cccgggcaac cactgatcct tttactg
<210> 251
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1481-285
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1481-285.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1481-285.mis2
<400> 251
                                                                        47
tcccataacc tgttttgctt ctcgctctaa cctcaagatg gtataaa
<210> 252
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1484-101
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1484-101.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1484-101.mis2
 <400> 252
                                                                        47
aaaaagatca aatataagca tgtaactcct ctccttaaaa tctcagt
 <210> 253
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
 <222> 1..47
 <223> polymorphic fragment 99-1484-328
 <221> allele
 <222> 24
 <223> polymorphic base G
 <221> primer_bind
 <222> 1..23
```

```
<223> potential microsequencing oligo 99-1484-328.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1484-328.mis2
<400> 253
                                                                       47
ggacacgtgg tcatgaggag tttgaaggga ttcagttttc agatccc
<210> 254
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1485-251
<221> allele
<222> 24
<223> polymorphic base G
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1485-251.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1485-251.mis2
<400> 254
                                                                       47
gattgccttg atatatgctc ccagagaacc aagaatgtcc ccttttc
<210> 255
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1490-381
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-1490-381.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-1490-381.mis2
<400> 255
tgcacagtgg aaataccatg tcacggtacg ctactgtgca tctcttc
                                                                        47
<210> 256
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-1493-280
<221> allele
<222> 24
<223> polymorphic base A
 <221> primer_bind
 <222> 1..23
 <223> potential microsequencing oligo 99-1493-280.mis1
 <221> primer_bind
 <222> 25..47
 <223> complement potential microsequencing oligo 99-1493-280.mis2
```

```
<400> 256
                                                                        47
ggatgacaga gtattgttgg aggaatgggg tttggctgct tgttttt
<210> 257
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-151-94
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-151-94.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 39-151-94.mis2
<400> 257
                                                                        47
attgagatca ttgataagga aatattctaa aatttcaaaa tctatat
<210> 258
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-211-291
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-211-291.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-211-291.mis2
<400> 258
                                                                        47
ctggttatat cagactgacc ttcatgtttt caacaggtca atgcctt
<210> 259
<211> 45
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..45
 <223> polymorphic fragment 99-213-37
<221> allele
 <222> 23
 <223> polymorphic base T
 <221> primer_bind
 <222> 1..22
 <223> potential microsequencing oligo 99-213-37.mis1
<221> primer_bind <222> 24..45
 <223> complement potential microsequencing oligo 99-213-37.mis2
 <400> 259
                                                                         45
 gtgcttccgg ctgcaggact gttggaggac tccagtgtct gacag
 <210> 260
 <211> 47
```

```
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-221-442
<221> allele
<222> 24
<223> polymorphic base A
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-221-442.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-221-442.mis2
<400> 260
                                                                        47
tgcctttgta gatatgcatg ggaattccat gacctagcca gacgaat
<210> 261
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 99-222-109
<221> allele
<222> 24
<223> polymorphic base C
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 99-222-109.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 99-222-109.mis2
<400> 261
                                                                        47
caggtgagga gtgctggatt ggccacgata tgaatttctt cagcagt
<210> 262
<211> 47
<212> DNA
<213> Homo Sapiens
<220>
<221> allele
<222> 1..47
<223> polymorphic fragment 4-14-107, variant version of SEQ ID185
<221> allele
<222> 24
<223> base G ; A in SEQ ID185
<221> primer_bind
<222> 1..23
<223> potential microsequencing oligo 4-14-107.mis1
<221> primer_bind
<222> 25..47
<223> complement potential microsequencing oligo 4-14-107.mis2
<400> 262
                                                                        47
ctaaacaacc accaaatgca tacggcaacc aggcaaatgc ctgatag
 <210> 263
 <211> 47
 <212> DNA
 <213> Homo Sapiens
 <220>
 <221> allele
```